



**UNITED STATES DEPARTMENT OF COMMERCE**  
**Economics and Statistics Administration**  
**U.S. Census Bureau**  
Washington, DC 20233-0001

This document was prepared by and for Census Bureau staff to aid in future research and planning, but the Census Bureau is making the document publicly available in order to share the information with as wide an audience as possible. Questions about the document should be directed to Kevin Deardorff at (301) 763-6033 or [kevin.e.deardorff@census.gov](mailto:kevin.e.deardorff@census.gov)

October 10, 2012

## **2010 CENSUS PLANNING MEMORANDA SERIES**

**No. 244**

MEMORANDUM FOR      The Distribution List

From:                      Burton Reist *[signed]*  
Acting Chief, Decennial Management Division

Subject:                    2010 Census Effectiveness of Unduplication Evaluation Report

Attached is the 2010 Census Effectiveness of Unduplication Evaluation Report. The Quality Process for the 2010 Census Evaluations, Experiments, and Assessments was applied to the methodology development, specifications, software development, analysis, and documentation of the analysis and results, as necessary.

If you have questions about this report, please contact Sarah Heimel at (301) 763-9297 or Ryan King at (301) 763-4774.

Attachment

September 25, 2012

# **2010 Census Effectiveness of Unduplication Evaluation Report**

U.S. Census Bureau standards and quality process procedures were applied throughout the creation of this report.

Sarah Heimel and Ryan King

---

Decennial Statistical Studies Division  
Center for Survey Measurement



**This page is intentionally left blank.**

## TABLE OF CONTENTS

1.	Introduction.....	1
1.1	Scope .....	1
1.2	Intended Audience.....	1
2.	Background .....	1
2.1	Identifying Duplicate Persons .....	1
2.2	Operationalizing Unduplication .....	3
2.3	Census 2000 .....	4
2.4	2004 Census Test .....	6
2.5	2005 National Census Test.....	7
2.6	2006 Census Test .....	8
2.7	2008 Census Dress Rehearsal .....	9
2.8	2010 Census .....	11
3.	Methodology .....	15
3.1	Research Components .....	15
3.1.1	Conduct the regular 2010 CFU interview with a sample of housing units containing potentially duplicated persons.....	15
3.1.2	When applicable, utilize experimental questions at the end of the CFU interview.	16
3.1.3	Conduct in-depth Qualitative Interviewing (QI) on a variety of duplicates. ....	17
3.2	Research Questions .....	18
3.2.1	Universe of Duplication Cases.....	18
3.2.2	CFU Sample.....	19
3.2.3	Experimental CFU Questions (Mod Q Cases).....	19
3.2.4	Qualitative Interviews .....	19
3.3	Sample Design.....	19
3.4	Sample Weighting.....	21
3.5	Data Sources.....	21
3.5.1	Duplicate Person Identification (DPI) Files.....	21
3.5.2	2010 Decennial Response File (DRF) .....	21
3.5.3	Unduplication Analysis Files .....	21
3.5.4	CFU Assessment Files .....	21
3.5.5	Address File .....	21
4.	Limitations .....	22
4.1	Incomplete Universe of Potential Duplicates .....	22

4.2	Mod Q Cases .....	22
4.3	Large Clusters .....	22
4.4	Recall Bias.....	22
4.5	Geographic Descriptors.....	23
5.	Results.....	23
5.1	Characteristics of the Unduplication Universe.....	23
5.1.1	Overview.....	24
5.1.2	Geographic Distance.....	25
5.1.3	Form Type.....	28
5.1.4	Number of Person Links .....	31
5.1.5	Whole and Partial Households.....	32
5.1.6	Phone Match .....	34
5.1.7	GQ Type.....	37
5.1.8	United States and Puerto Rico .....	40
5.1.9	Overcount Question .....	41
5.1.10	Demographic Characteristics of Linked Persons .....	44
5.2	Sampled Cases for CFU .....	58
5.2.1	Sampling Overview .....	58
5.2.2	Case Completion.....	62
5.2.3	Complex Situations.....	70
5.2.4	Resolving Duplication .....	83
5.3	Experimental CFU Questions (Mod Q Cases).....	95
5.3.1	Universe for Mod Q.....	96
5.3.2	Mod Q Overcount .....	98
5.4	Qualitative and Cognitive Research.....	104
5.4.1	Objective .....	104
5.4.2	Methodology .....	105
5.4.3	Findings from the Cognitive Interviewing with Suspected Duplicates .....	109
5.4.4	Findings from the Qualitative Interviewing with Suspected Duplicates .....	126
5.4.5	Dependent Interviews .....	129
5.4.6	Memory Issues for Respondents.....	133
5.4.7	Privacy and Sensitivity .....	133
6.	Related Census Assessments .....	133
7.	Conclusions and Recommendations .....	134

8. Acknowledgements .....	142
9. References .....	143
APPENDIX A – 2010 Census Residence Rule and Residence Situations .....	147
APPENDIX B – Person Counts by Form Type by Wave.....	154

## LIST OF TABLES

Table 1. Number of Response and Person Links by Link Type .....	24
Table 2: Geographic Relationship of Links .....	25
Table 3. Distribution of Person Links by Geography and Link Type .....	26
Table 4. Distribution of Response Links by Geography and Link Type .....	26
Table 5: Distance Between ZIP Codes of Response Links.....	27
Table 6. Distance Between ZIP Codes of Response Links by Geography .....	28
Table 7. Form Types Linked at the Person Level .....	29
Table 8. Form Types Linked at the Person Level, by Geography .....	30
Table 9: Number of Person Links (Per Household) Between HU-HU links .....	31
Table 10. Number of Person Links Between HU-HU Links, By Geographic Proximity.....	32
Table 11: Whole/Partial Status by Type of Link for Response Links .....	33
Table 12. Geographic Distribution of Whole/Partial Status by Response Links.....	34
Table 13: Frequency of Phone Numbers Matching (HU-HU).....	35
Table 14. Phone Matches at the Housing Link Level by Geography .....	36
Table 15. Phone Matches at the Housing Link Level by Form Type .....	36
Table 16. Phone Matches at the Housing Link Level by Whole/Partial Status.....	37
Table 17: Type of GQ .....	38
Table 18. Number and Percent of GQ Types by Geography .....	39
Table 19: Frequency of Matching Between Stateside Addresses and Puerto Rico Addresses....	40
Table 20. US and Puerto Rico Links at the Person Link Level, by Geography .....	41
Table 21: Overcount Indication Provided By HU to HU Links .....	42
Table 22: Overcount Indication Provided By HU to GQ Links .....	42
Table 23: Presence of Overcount Mark for HU to HU Person Links, By Geography .....	43
Table 24: Presence of Overcount Mark for HU to GQ Person Links, By Geography .....	43
Table 25. Age Category Match Status By Geography.....	45
Table 26. Age of Duplicates by Geography.....	46
Table 27. Age of Duplicates Compared to Age of All Enumerated Persons.....	47
Table 28. Hispanic Origin Category Match Status By Geography.....	48
Table 29. Hispanic Origin of Duplicates by Geography.....	49
Table 30. Hispanic Origin of Duplicates Compared to All Enumerated Persons.....	50
Table 31. Race Category Match Status By Geography .....	51
Table 32. Race of Duplicates by Geography .....	52
Table 33. Race of Duplicates Compared to All Enumerated Persons .....	54
Table 34. Sex Match Status By Geography .....	55
Table 35. Sex of Duplicates by Geography .....	56
Table 36. Sex of Duplicates Compared to All Enumerated Persons .....	57
Table 37. CFU Eligible and Sent Workload .....	60
Table 38. Response Links Sent to CFU, by Type of Link (unweighted).....	60
Table 39: Percent of CFU-Eligible Response Links Sent To CFU (unweighted) .....	61
Table 40: Completion Rate by Strata, HU to HU Response Link (unweighted).....	63
Table 41: Completion Rate by Strata, HU to GQ Response Link (unweighted).....	64
Table 42: Completed Interviews by Type of Person Link (unweighted).....	64
Table 43: Completed Interviews by Strata for HU-HU Person Links (unweighted).....	65
Table 44: Number of Unweighted and Weighted Person Links From Completed Interviews.....	66

Table 45: Completion Rates by Type of Person Link (Weighted) .....	67
Table 46: Completed Person Links by Stratum, for HU-HU links (Weighted).....	68
Table 47: Completion Percentage from HU-HU Person Links, by Phone Number Match .....	69
Table 48: Distribution of Linked Form Types, by CFU Universe and Overall .....	69
Table 49: Rate of Complex Living Situations Being Mentioned in Person Links .....	71
Table 50: Rate of Complex Living Situations Being Mentioned for Person Links, by Stratum ..	72
Table 51: Rate of Complex Living Situations Being Mentioned for HU-HU Person Links When Both Sides Completed a CFU Interview, by Stratum .....	73
Table 52: Rate of Complex Living Situations Being Mentioned for HU-HU Person Links When One Side Completes a CFU Interview, by Stratum .....	74
Table 53: Rate of Complex Living Situations Being Mentioned for HU-GQ Person Links, by Stratum.....	75
Table 54: Frequency that the CFU Complex Situation Mentioned Agreed with the Overcount Strata .....	77
Table 55: Complex Situations for Other Above the Line and Below the Line Overcount Cases – HU to HU Person Links .....	78
Table 56: Complex Situations for Above the Line and Below the Line Overcount Cases – HU to GQ Person Links .....	79
Table 57: Type of Complex Situations for Unduplication-Only HU-HU Links .....	80
Table 58: Type of Complex Situations for Unduplication-Only HU-GQ Links .....	82
Table 59: Deletion Rate of Person Links .....	84
Table 60: Deletion Rates by Stratum – Entire Universe of HU to HU Person Links .....	85
Table 61: Deletion Rates by Stratum – Entire Universe of HU to GQ Links .....	86
Table 62: Deletion Rate for Person Links that Mentioned a Complex Situation, by Type of Link .....	87
Table 63: Deletion Rate for HU-HU Person Links that Mentioned a Complex Situation, by Stratum.....	88
Table 64: Deletion Rate for HU-GQ Person Links that Mentioned a Complex Situation, by Stratum.....	89
Table 65: CFU Residence Status for HU-HU Person Links .....	90
Table 66: CFU Residence Status for HU side of HU-GQ Person Links .....	90
Table 67: Deletion Rate of Person Links by Phone Match Status .....	91
Table 68: Deletion Rate of Person Links by Geography .....	92
Table 69: Deletion Rate of Person Links by Age .....	93
Table 70: Deletion Rate of Person Links by Sex .....	94
Table 71: Deletion Rate of Person Links by Hispanic Origin .....	94
Table 72: Deletion Rate of Person Links by Race .....	95
Table 73: Eligible and Completed Mod Q Persons .....	97
Table 74: Fate of Duplicate Persons Eligible for Mod Q, by Initial Overcount Categories .....	98
Table 75: Reason Given In Mod Q for Marking the Initial Overcount Question .....	99
Table 76: Reason Given In Mod Q for Marking the Initial Overcount Question, by Initial Overcount Mark .....	100
Table 77: Responses to the MOTT Question in Mod Q for Persons Away from the CFU Address .....	102
Table 78: Responses to the MOTT Question in Mod Q for Persons Away from the CFU Address, by Overcount Mark .....	103



Table 79: Targeted Questions for TCFU, By Age Classification .....	106
Table 80: Differences between the CFU and TCFU interviews .....	108
Table 81: Distribution of Cases Cognitively Interviewed .....	110
Table 82: Distribution of Cases Cognitively Interviewed, by Respondent .....	111
Table 83: Success Rate in Cognitive Interviews, by Case Type.....	113
Table 84: Success Rate in Cognitive Interviews, Type 1 Cases .....	114
Table 85: Success Rate in Cognitive Interviews, Type 2 Cases .....	116
Table 86: Success Rate in Cognitive Interviews, Type 3 Cases .....	117
Table 87: For Verified Duplications, Reasons for Duplication by Type of Case.....	118
Table 88: Success Rate in Cognitive Interviews, by Respondent.....	118
Table 89: Success Rate in Cognitive Interviews, by Number of Duplicates .....	119
Table 90: Success Rate in Cognitive Interviews, by Household Match .....	120
Table 91: Success Rate in Cognitive Interviews, by Household Match and Number of Duplicates .....	120
Table 92: Frequency of Address Completeness.....	121
Table 93: Number of Complete and Partial Addresses Provided by TCFU Questions .....	122
Table 94: Success Rate in Qualitative Interviews.....	126
Table 95: Reasons for Duplication in Qualitative Interviews.....	127

## LIST OF FIGURES

Figure 1: Undercount Coverage Probe .....	12
Figure 2: Overcount Coverage Probe .....	13

## EXECUTIVE SUMMARY

Counting each person once, only once, and in the right place is the foundation of the decennial census. Oftentimes though, people have multiple places where they spend time and so could be enumerated at more than one place, creating duplication in the census.

The Census Bureau has developed computer-matching algorithms to match the census universe against itself and thus identify potentially duplicated persons. The algorithms use characteristics such as first name, last name, middle initial, age, date of birth, phone number, and geographic distance to match people. Each time a person record is matched to another person record, it is given a score that reflects the strength of the match. The scores are then ranked and the matches are reviewed to establish a cutoff point. All matches with scores above the cutoff are reliably identified as duplicate person records. Cutoffs are set very high during the review to minimize false matches being incorrectly classified as duplicates. Followup operations were expensive in the 2010 Census so resources could not be wasted on false matches. The computer-matching process only identifies potential duplicates; no individuals are removed from the census during this process. Although extensive research has been done to ensure that chance agreements of name and date of birth are not classified as matches, and while the cutoffs are high, there is still the possibility that persons matched as potential duplicates are not actual duplicates. On the other hand, computer matching will fail to identify some duplicates because of inaccurate or missing data.

The computer-matching algorithm identifies an association of one person to another, called a “link.” The Census Bureau is interested both in the individuals who are linked and in the housing units occupied by those individuals. Two linked people are considered to be a “person link.” The housing units involved in each person link are known as “housing unit links.” The census questionnaires that enumerate the linked people are known as “response links.”

The universe of all housing unit returns in the 2010 Census was matched against itself to identify people who may have been duplicated. Group Quarters returns were also included and compared to housing unit returns. For the scope of this research however, census returns were only included if they were data captured by the end of July 2010 and were in scope for the Coverage Followup operation.

The purpose of this evaluation is:

- To document the universe of duplication cases identified in the 2010 Census,
- To document the results of duplication cases sent to the Coverage Followup operation,
- To document the results of the experimental questions asked to a subset of duplicated persons at the end of the Coverage Followup interview, and,
- To convey the results of the cognitive and qualitative interviews conducted with duplication cases.

### **What did the universe of duplication cases look like?**

There were 4,711,560 response links identified from 2010 Census returns that contained at least one potentially duplicated person. This resulted in a total of 7,454,171 person links, or

potentially duplicated persons. Housing unit to housing unit person links accounted for 88.6 percent of all potential person duplication. Housing unit to Group Quarters links accounted for the remaining 11.4 percent of potential person duplication.

Within the housing unit to Group Quarters links, over half (51.7 percent) were between a housing unit and student housing, such as college dormitories.

All links are classified by the geographic proximity of the two addresses; within-block, within-surrounding-blocks, within-county, within-state, or across state lines. Person duplication occurred most frequently within-block (33.5 percent of the time) followed by within-county (28.0 percent).

Links were also categorized by the response mode used. Questionnaires were categorized as either being a Group Quarters questionnaire, being a mailed back housing unit form, or being an enumerator-completed housing unit form. Person duplication occurred most frequently (54.5 percent of the time) between one Mailout/Mailback questionnaire and one Enumerator questionnaire, followed by duplication 26.8 percent of the time between two Mailout/Mailback questionnaires from two different housing units (two distinct Master Address File Identification Numbers).

In the majority of housing unit to housing unit response links (58.5 percent), only one person was suspected to be duplicated between the two housing units. Two people were suspected to be duplicated between the two housing units in 23.6 percent of links.

If two housing unit responses provided the same phone number on their return, that increased the confidence of persons across each response being a match. Almost one-quarter (24.6 percent) of all housing unit to housing unit response links had a matching phone number.

The overcount question on the initial census enumeration is intended to capture possible erroneous enumerations. While not used during the computer-matching process to identify duplicates, frequency of positive indicators to the overcount question was examined with other characteristics of suspected duplicates. Over 40 percent (41.3 percent) of suspected housing unit to housing unit duplicates marked the overcount question on at least one of the two linked responses and almost sixty percent (58.7 percent) of suspected housing unit to Group Quarters duplicates marked the overcount question on the housing unit response.

### **How successful were the cases sent to production Coverage Followup?**

The Coverage Followup operation utilized a personal telephone call to housing units, after a census questionnaire had been received for that housing unit, in order to resolve different types of coverage issues, including:

- returns with more people reported in the household population count box than were able to be fully detailed on the census return person panels<sup>1</sup>,
- returns that had a count discrepancy,

---

<sup>1</sup> These were called “large households.” For instance, this occurred on a mailback form that had more than six people enumerated.

- returns that flagged the undercount or overcount question, and
- returns flagged by Administrative Records processing

The Coverage Followup interview used an in-depth questionnaire asking about different types of living situations in order to resolve person-level coverage issues.

During the mid-decade testing phase, it was determined that the Coverage Followup interview was not as successful in resolving duplication as it was in resolving other coverage issues. For instance, almost 80 percent of the potential duplicates from the 2008 Dress Rehearsal were not resolved during CFU and over 90 percent of the duplicates from the 2006 Census Test were not resolved during CFU. Due to this lack of success and budget constraints, housing units with persons identified as potential duplicates were not eligible for production followup as part of the 2010 Census. However, a sample of duplication cases was selected and sent to Coverage Followup in order to assess how successful the Coverage Followup interview was in a decennial Census environment. The results of these sampled cases did not impact the final Census results; however, if the case contained a duplicate person and was already in the production flow for another reason (such as having a count discrepancy), then the results would impact the final Census results. All results presented in this report ignore the production or sampled nature of cases. The sample was stratified using the presence of an overcount marking on the initial census return, the geographic proximity of the links, and whether the links were between two housing units or between a housing unit and a Group Quarters.

There were 469,768 response links sent to Coverage Followup (15.0 percent of all 3,137,840 within county, within state, and across state response links that had been identified). We completed a Coverage Followup interview with at least one housing unit in 350,757 response links, or 74.7 percent of the time. Those 350,757 response links represent 424,806 unweighted person links and 2,209,562 weighted person links. The success of the Coverage Followup interview was determined by investigating how often respondents mentioned a complex living situation for a suspected duplicated person and how often a suspected duplicated person was marked for removal from a housing unit roster based on the interview results, thus indicating that the duplication would have been resolved.

Only 58.7 percent of person links had at least one side mention a complex living situation in the Coverage Followup interview. Some duplication cases that were sent to Coverage Followup had marked the overcount question on the initial return, indicating a possible erroneous enumeration, while other duplication cases sent to Coverage Followup had not marked the overcount question. Person links that had marked the overcount college box and contained a suspected duplicate person were the most likely to mention a complex living situation in the Coverage Followup interview. Person links that had not marked an overcount box on either side of the link were less likely to mention a complex living situation in Coverage Followup than links that had marked the overcount box. For instance, within-county links between two housing units when neither marked the overcount box only mentioned a complex living situation in Coverage Followup 20.2 percent of the time. However, 85.6 percent of links between two housing units where at least one side had marked the college overcount category subsequently mentioned a complex living situation in Coverage Followup.

The Coverage Followup interview was successful in resolving duplication for 31.2 percent of all the person links. However, a complex living situation had to have been mentioned during the Coverage Followup interview in order for a person to be removed. Of the person links that had mentioned a complex living situation in the Coverage Followup interview, 53.2 percent removed someone from the roster. This figure was higher for links between a housing unit and a Group Quarters, which deleted someone 73.7 percent of the time when a complex living situation had been mentioned.

Of the links between two housing units where an indication had been provided on the initial census return of a seasonal or second home, 44.0 percent of the links deleted someone as a result of Coverage Followup. That rate increased to 57.8 percent of such links that mentioned a complex living situation in the followup interview. This result is promising for resolving duplication involving seasonal residences in a true census environment.

However, the majority of duplication comes from persons who do not mark an overcount reason. For person links that were within the same county and neither side marked an overcount reason, only 7.0 percent had a person deleted as a result of Coverage Followup. For person links within the same state when neither side marked an overcount reason, 10.1 percent had someone deleted as a result of Coverage Followup. For person links across state lines when neither side marked an overcount reason, only 6.9 percent had someone deleted as a result of Coverage Followup.

### **What information was obtained about the living situation of duplicates from the cases that were sent to Mod Q?**

If a person marked an overcount or undercount reason that was not resolved in the Coverage Followup interview, then a sample of cases were eligible for an experimental series of questions at the end of the interview (called ‘Mod Q’) to probe the respondent about why they had indicated an overcount or undercount reason on their initial response. The responses to these experimental questions would not impact the final Census results.

There were 125,370 persons identified as potential duplicates who were eligible for the questions in Mod Q. There were 29,548 persons who completed the overcount series of Mod Q questions, with 42.5 percent of those persons being eligible because they had initially marked the ‘Another reason’ overcount category.

The majority of persons, 57.0 percent, provided an open-ended response for why they marked the overcount category. These responses were not coded in time to be included in this evaluation, but should be investigated once they are available.

There were 9,694 respondents who indicated that they marked the overcount category because the person in question:

- stayed away in March/April 2010,
- stayed away sometime in 2010,
- stayed away briefly, or
- stayed at another address for a different reason.

We then asked those respondents where the person in question spent most of their time (the 2010 Census address, the other address, or both addresses equally). Over half of the respondents (52.2 percent) indicated that the person spent most of their time at the Census address (the address being asked about in the Coverage Followup interview). Only 23.9 percent of persons indicated the person spent most of their time at the other place, while 22.3 percent indicated that the person split their time equally between addresses. Since the results of Mod Q did not impact the final Census count, there was no question in this experimental series about a person's location on April 1.

### **What results were obtained from the cases that completed a Qualitative or Cognitive Interview?**

A study was commissioned by the Census Bureau and conducted by RTI International and Research Support Services using 2010 Census data to investigate the characteristics and questions of a followup interview that would prompt respondents to provide the information necessary to confirm and resolve person duplication without anyone feeling that privacy or confidentiality had been violated (Peytcheva et al, 2011). Two substudies, one qualitative and one cognitive, were launched simultaneously to address the outlined objectives.

A total of 50 qualitative interviews and 226 cognitive interviews were completed, all with cases that had not marked the overcount question on either initial census response. Cases were classified into three distinct categories:

- If a person (or persons) was counted on two distinct housing unit returns, and the phone number was the same on each return, it was considered a Type 1 Case.
- If a person (or persons) was counted on two distinct housing unit returns, and the phone numbers were different, it was considered a Type 2 Case.
- If a person was counted in a Group Quarters facility as well as in a housing unit, it was considered a Type 3 Case.

A scripted interview was used in the cognitive study (known as the Targeted Coverage Followup instrument), though the interview contained different questions depending on whether the case was a Type 1, Type 2, or Type 3 case. Targeted Coverage Followup interviews were to be distributed across the combination of Type 1, Type 2, and Type 3 cases. Within Type 1 and Type 2 cases, interviews were to be distributed across both whole household matches and partial household matches, and with a variety of ages reflected in the duplicated persons. Within Type 3 cases, interviews were to be distributed across different types of Group Quarters in which persons could be duplicated.

The qualitative study used a flexible style of interviewing. Researchers were also to attempt to interview both sides of a link and compare the answers. These interviews with both sides of a link were referred to as dependent interviews. After each interview, an assessment was made to whether enough information had been provided to verify the duplication. Non-verified duplicates were cases where the research staff believed it was, or could have been, the same person listed on each side of the link, but the respondent did not provide enough information in the interview to confirm that. False matches, or suspected false matches, were cases that were reviewed by interviewers and Census Bureau staff and (based on the information obtained in the

interview as well as the information provided on the initial census questionnaires) were thought to represent two different people on each side of the link, instead of an actual duplication.

From the cognitive Targeted Coverage Followup interviews, Type 1 cases verified the duplication 97.6 percent of the time, with the other 2.4 percent of cases unverified. Type 3 cases verified the duplication 69.5 percent of the time, with 16.9 percent non-verified and 13.6 percent either a false match or a suspected false match. Type 2 cases verified the duplication 58.4 percent of the time, with 21.6 percent non-verified and 20.0 percent classified as either false matches or suspected false matches. Type 1 cases, because they had the same telephone number on each return, were provided with both addresses during the interview (a change from the standard procedure), which was integral to achieving such high success rates.

Moving and having other property were the two main reasons for duplication among Type 1 cases. Those were also two of the three most common reasons for duplication among the Type 2 cases, in addition to staying at a relative's house.

Address information was collected about the living situations described in the Targeted Coverage Followup interviews. Complete addresses (meaning that at least the house number, street name, city, and state were provided) were collected 90.5 percent of the time when a respondent reported a move, 100.0 percent of the time when the respondent reported owning another residence, and 72.2 percent of the time when a seasonal or vacation home was reported. Partial address information was provided 84.6 percent of the time when a stay with relatives was reported, 78.6 percent of the time when a stay with a significant other was reported, and 75.0 percent of the time for stays with a brother or sister.

The Targeted Coverage Followup instrument was designed to provide both addresses to respondents when a phone match existed (Type 1 cases). This strategy seemed to work well, and only one participant was confused after receiving both addresses at the beginning of the interview. In all other Type 1 cases, presenting both addresses was perceived well by respondents and no one expressed any privacy concerns.

Of the 50 qualitative interviews, 35 (70.0 percent) verified the duplication within the interview. The most common reason for duplication in the qualitative interviews was a custody situation, with moving as the second most common reason. After seeing an initial preponderance of movers, Census Bureau staff had requested a focus on cases with duplication suspected to result from custody or second homes. Movers were attempted to be screened out during recruiting.

Seven dependent interviews were successfully completed within this study. Within these dependent interviews, the cases are "verified", in the sense that the address of the other place is revealed in the interviews on both sides and the identity of the duplicate is agreed upon. Beyond that, however, great differences emerge in the account of the circumstances surrounding the time spent in the two places, despite the apparent openness of the respondents on their own in these situations. Relying on both of these reports jointly, it would not be possible to determine the proper enumeration for the duplicated person in six of the seven situations.

## Conclusion and Recommendations

Leading up to the 2010 Census, the Census Bureau tested and implemented the most extensive efforts to date with the goal of resolving duplication. This evaluation has shown that there are certain duplicate situations that we were successful at resolving, but that there are other situations, despite our best efforts, where we fell short in resolving duplication. Persons who moved or had a seasonal home were cooperative in both CFU and the qualitative interviews, providing addresses and describing their living situation. Respondents for children in custody situations often mentioned that living situation, but it was often difficult to resolve at which one address they should be counted. Duplications between housing units and Group Quarters were often resolved in both CFU and the qualitative interviews. Respondents in the qualitative interviews who had provided the same phone number on two returns were cooperative and able to explain and resolve the duplication when presented simultaneously with both addresses. These cases show promise for resolution in 2020, if we continue to use a similar followup procedure. One-third of all duplication was classified as occurring within the same block, which could potentially be resolved in the future through continued improvements to the Master Address File or the use of automated questionnaires that confirm the address of the respondent, to reduce housing unit mix-ups.

Moving towards the 2020 Census, there is much room for improvement in the resolution of duplication and the production of a more accurate census. In order to improve the effectiveness of unduplication efforts in future censuses, we recommend the following actions:

- **Gain more knowledge on duplication in the 2010 Census.**
  - Repeat the analysis presented in Section 5.1 of this report using the complete universe of census returns.
  - Research large clusters (links involving more than two Master Address File Identification numbers).
  - Study address-level duplication more, including a further analysis of the 2010 Field Verification results.
  - Conduct additional analysis with address data obtained in the Coverage Followup interview.
  - Conduct additional research using detailed responses from Coverage Followup when interviews were completed with both sides of a link.
  - Research the open-ended answers to the first question in Mod Q (asking why the person was living somewhere else).
- **Attempt to reduce false matches and capture more true matches in future matching.**
  - Explore a way to include relationship status and family composition in the probabilistic weighting of a match, especially when only one person in a household is thought to be duplicated.
  - Consider clarifications to the current categorization of geographic proximity (within block, etc.) or consider using additional variables to describe the actual distance that two addresses are from each other (such as latitude and longitude, or ZIP codes).
- **Consider the implications and possibilities with automation for identification and resolution of duplicates in 2020.**



- Collect alternate address information on the initial enumeration when respondents indicate they have another place where they live or stay.
- Once collected, utilize alternative address information to identify duplicates.
- **Seek to improve current practices that influence duplication.**
  - Continue to research improvements to the overcount question and probes, to capture as many erroneous enumerations as possible.
  - Continue to refine the targeted followup approach in utilizing existing data about a duplication to make any followup contacts more efficient and successful.
  - Continue to improve the address list development process to minimize address duplication.
  - Improve the identification of Group Quarters as Group Quarters.
  - Expand Geography Division's automated matching process to utilize building and Group Quarters names when matching addresses to the Master Address File.
  - Consider a clerical operation during production to resolve lingering duplication cases in the future.
  - Communicate with the public about followup efforts.
- **Review current Census Bureau policies related to duplication.**
  - Consider resolving the duplication according to the residence rule without any additional contacts to the housing unit if:
    - Alternative address information can be collected from the initial returns, processed automatically, and utilized to confirm duplication between a housing unit and a Group Quarters, or
    - Alternative address information as well as sufficient information on living patterns can be collected initially for certain Housing Unit to Housing Unit duplicate situations (such as movers or seasonal residences)
  - Review the legal and political implications of counting persons at multiple residences.
  - Consider how Administrative Records could be utilized either to confirm a suspected duplication or aid in the resolution of where a duplicated person should be counted.
  - Review the policies for contacting suspected duplicates to ensure that no violations of privacy or confidentiality occur in any new methods of resolving duplication.

## **1. Introduction**

Counting each person once, only once, and in the right place is the foundation of the decennial census. Frequently though, people have multiple places where they spend time and so could be enumerated at more than one place. Additionally, a living quarter could be duplicated on our Master Address File (MAF). These situations could potentially create duplication in the census. The Census Bureau needs to be able to identify duplication and then resolve it by determining which census records to keep and which to remove. However, efforts to resolve duplication in the last decade have been disappointingly unproductive; this evaluation seeks to advance our knowledge of why person duplication occurs and how it can be successfully resolved in the future.

### **1.1 Scope**

The purpose of the Effectiveness of Unduplication Evaluation is to document the results and major findings of the efforts to identify and resolve duplication in the 2010 Census. This includes topics such as characteristics of the universe of identified duplicates, the success of cases worked by the Coverage Followup operation (CFU), the outcome of experimental questions implemented for the first time at the end of the CFU interview, and a qualitative component involving in-person interviews with suspected duplicates. This evaluation will inform stakeholders and decision makers of recommended changes or improvements for future censuses.

### **1.2 Intended Audience**

This document assumes that the reader has at least a basic understanding of the process of identifying duplicate persons, the CFU operation, and the Field Verification (FV) operation. The goal is to use this document to help research, planning, and development teams in planning for the 2020 Census. A basic overview of the duplicate person identification (DPI) process, the CFU operation, and the FV operation is provided in this document. For more information on the duplicate person identification process, please refer to the Decennial Statistical Studies Division 2010 Decennial Census Memorandum Series I-01 (Lynch, 2009), I-02R1 (Frank, Ikeda and Porter, 2011a), and I-03R1 (Frank, Ikeda and Porter, 2011b). For more information on the CFU operation, please refer to the 2010 Census Coverage Followup Assessment (Govern, Coombs and Glorioso, 2012). For more information on the FV operation, please refer to the 2010 Census Field Verification Operational Assessment (McPhillips, 2012).

## **2. Background**

### **2.1 Identifying Duplicate Persons**

People can be duplicated in the census for reasons related either to their living situation (called person-level duplication) or for reasons related to the physical address at which they live (called housing-level duplication).

In person-level duplication, a person may have been included on more than one questionnaire for reasons such as, but not limited to:

- Joint custody situations,
- Enrollment in college,
- Ownership of multiple residences, or,
- Other reasons that led to part-time residency situations.

Persons who spend time at more than one place and consequently may be enumerated more than once are considered to have complex living situations.

In housing-level duplication, a person (or household of people) may have been included on more than one questionnaire either because:

- A housing unit appeared more than once on the Census Bureau's master address list and so received multiple questionnaires, all of which were completed either by the household or by a Census enumerator. An address issue like this has the potential to be resolved in the FV operation.
- A questionnaire was misdelivered, called form misdelivery. If the post office incorrectly delivered a questionnaire, neighbors may have received each other's census form. For example, Apartment A might have received Apartment B's questionnaire and vice versa. In the best-case scenario, both apartments A and B returned their questionnaires by mail and the Census Bureau would never know the difference – all the people were enumerated in the correct block, even though they were counted in the wrong housing units. If, however, only one unit returned the questionnaire, the Census Bureau would followup in person with the nonresponding housing unit to get the missing questionnaire. The barcodes on the forms indicated which housing units returned the questionnaire, but in the form misdelivery case it would have appeared that a housing unit had returned the questionnaire when they actually had not. The Census Bureau may then have conducted an interview with the unit that had already completed a questionnaire (though that initial questionnaire was associated with a different housing unit). Procedures existed that allowed nonresponse followup interviewers to correct these form misdelivery situations in the field, but if for some reason the interviewers did not or could not correct the problem, then duplication of persons will occur. This form misdelivery has resulted in an apartment mix-up.

The Census Bureau has developed computer-matching algorithms that match the census universe against itself and identify potentially duplicated persons. The algorithms use characteristics such as first name, last name, middle initial, age, date of birth, phone number, and geographic distance to match people. The process involves multiple passes of the system where the matching parameters and constraints are varied for each pass. Each time a person record is matched to another person record, it is given a score that reflects the strength of the match. The scores are then ranked and the matches are reviewed to establish a cutoff point. Cutoffs are set very high during the review to establish a high level of certainty that only true duplicates and not false matches are identified. All matches with scores above the cutoff are considered to be duplicate person records.

Although extensive research has been done to ensure that chance agreements of name and date of birth are not classified as matches, and while the cutoffs are high, there is still the possibility that

persons matched as potential duplicates are not actual duplicates (Fay 2002, Fay 2004, Ikeda and Porter 2007, Ikeda and Porter 2008, Yancey and Winkler 2002, and Yancey 2007). Also, computer matching will fail to identify some duplicates because of inaccurate or missing names or dates of birth. Generally, the Census Bureau prefers to be conservative and not identify some duplicated people rather than identify false matches. This conservative approach allows our limited resources to focus on links we suspect to be true duplicates.

The computer-matching algorithm identifies an association of one person to another, called a “link.” The Census Bureau is interested both in the individuals who are linked and in the housing units occupied by those individuals. Two linked person records are considered to be a “person link.” The housing units (HUs) involved in each person link are known as “housing unit links.” The census questionnaires (or responses) that enumerate the linked people are known as “response links.”

Computer matching can also bring together more than two records. For example, if a college student is counted in a dormitory through the Group Quarters Enumeration (GQE) operation and both of the divorced parents include the student on their housing unit census returns, three census responses would be linked during matching. Such groups of three or more responses are called “large clusters.”

## 2.2 Operationalizing Unduplication

Each person link found by the computer matching is identified and classified to help decide the best way to resolve the duplication. The first classification is based on whether the person records were found in housing units or in Group Quarters (GQs). Person links can either be found in two distinct housing units, called HU-HU matches, or between a housing unit and a GQ, called HU-GQ matches. No matching is done to identify person links between two different GQs because the enumeration that takes place at GQs is the final enumeration outcome by design of the operation and there is no mechanism in place to resolve such a duplication.

The HU-HU matches are also classified based on the distance between the housing units. There are five levels at which links can be geographically associated. The levels below are mutually exclusive and are listed in order of precedence:

- a. Within the same block
- b. Within two different blocks, which are adjacent to each other, called surrounding blocks
- c. Within the same county
- d. Within the same state
- e. Outside of the state

Duplication can be resolved by determining at which one location each individual should be counted, but operationalizing that goal is complex. There is usually insufficient information on the initial census enumeration to apply the census residence rule and determine the correct

location for a person duplicated across two housing units.<sup>2</sup> Followup interviews need to be conducted in order to learn more about the duplicated person's living situation and make an informed decision on whether each person record on a census questionnaire accurately reflects their census residency.

The followup interview needs to be conducted with a knowledgeable respondent at each address to obtain accurate information. The Census Bureau must conduct the interview without violating confidentiality, sensitivity, and Title 13 concerns. Most notably, this means that the respondent cannot be told that the duplicated person was listed on a second census return. The address or general location of the suspected duplicate enumeration also cannot be revealed to the respondent. In order to resolve the duplication, the respondent needs to voluntarily mention another address where the duplicated person lives or stays and then also describe how much time the person spends at that location.

## 2.3 Census 2000

The Census 2000 address list development process was susceptible to including duplicate housing units. A process was implemented to identify and remove from the census duplicate housing units that still remained on the decennial file after all data collection activities had been completed. Both address-level and person-level matching were performed to identify potential duplicates.

An ad hoc operation called the Housing Unit Duplication Operation (HUDO) was mounted to research and eliminate certain categories of duplicate housing units from final Census 2000 counts. Approximately 1.4 million housing units were deleted as a result of HUDO (Nash, 2000b). The operation also resulted in the removal of approximately 3.6 million enumerated persons within these HUs from the census (Nash, 2000b). No Census 2000 operation directly addressed or resolved problems stemming from person-level duplication; the Coverage Edit Followup Operation (Sheppard, 2003) might have resolved some duplication that existed in count discrepancy cases<sup>3</sup>, but those cases were not targeted for followup because of the duplication.

The HUDO operation relied in part on computer matching of names and birth dates to identify potential duplicates. Census 2000 was the first census to incorporate the computer capture of census information using Optical Character Recognition as an integral part of the processing. This meant computer matching of person records could occur soon after questionnaires were returned, to expediently identify and address suspected duplicates in the census.

After Census 2000 had concluded, a working group was established to address issues around duplication in the census. One of the group's proposals was to conduct a study to examine the effectiveness of telephoning sets of possible person duplicates found by computer matching and

---

<sup>2</sup> In the majority of cases where a person is duplicated between a housing unit and a Group Quarters (GQ), the person should be counted in the GQ according to the census residence rule (Lamas 2009). There are some exceptions to this, such as if the GQ is a religious group quarters or an in-patient hospice facility.

<sup>3</sup> Count discrepancies exist when the number of people reported in the population count box is different (either higher or lower) than the number of people detailed across the person panels.

then use followup techniques to resolve the residence status of the potential duplicates. The goal was to verify whether a person link was valid, identify why the person(s) were duplicated, and determine where the duplicate person(s) should have been counted on Census Day in 2000. This study was known as the Long Distance Duplicate (LDD) research (Smith, 2004) and was conducted in summer 2003 in the National Processing Center (NPC). The NPC analysts conducted the telephone followup using a modified version of what would become the 2004 Census Test Coverage Research Followup (CRFU) instrument.

To maintain confidentiality, the LDD study only asked about people captured in the census at the particular housing unit being called. The study never asked the respondent from one housing unit about people enumerated in the census at the other housing unit to which they had been linked. The content of the questionnaire included:

- Asking whether the interviewer had reached the correct household at the correct followup address and that the respondents had been living there on April 1st, 2000,
- Reminding the respondent of the people who were listed on the census form as of April 1, 2000, and
- Asking if any of those people could have stayed or lived anywhere else in 2000, including to attend college, live at another residence for a job, stay somewhere else for a joint custody situation, have a vacation home, stay with friends or relatives, or stay in a long-term care facility, military barracks, or correctional facility.

The universe of linked pairs for the LDD study was defined by a number of characteristics. The following situations made pairs ineligible for the study:

- Cases that were in either of the test sites for the 2004 Census Test were not eligible for the LDD study.
- If either return from a linked pair did not have a telephone number present, the pair was ineligible.
- Any links with a GQs return were ineligible.
- Links within the same county were also ineligible.

Results from the LDD study are limited, largely due to the time lag between the census enumeration and the followup interview (over three years) but also because the study was limited to no more than 500 interviews (Smith, 2004). High nonresponse was expected so a total of 736 linked pairs were sampled. The time lag led to a preponderance of disconnected phone numbers and difficulty in reaching knowledgeable respondents; 41.6 percent of the linked pairs were not interviewed because the phone numbers were disconnected or invalid (Smith, 2004). Thus, the LDD study completed interviews with 430 links. Half of those links (50.7 percent) were unable to be resolved due to language barriers, lack of a knowledgeable respondent, respondent refusals, or insufficient information. Of the 430 links, 32.8 percent confirmed a duplicated person. Duplication could not be confirmed for the remaining cases. The most common reasons why a person was duplicated, in order of frequency, were:

- Moving
- Visiting family or friends for an extended time
- Being enumerated at both a primary residence and seasonal home
- Having a joint custody situation
- Living at one residence for college and a second at the family home (Smith 2004).

Of the confirmed duplicates, 91.5 percent provided enough information that a decision could be made about where the duplicated person resided and should have been counted on Census Day (Smith 2004).

The LDD study hypothesized that by identifying duplicates with computer matching and then conducting some type of followup interview closer to census day, the true residence status could be ascertained for a number of duplicate persons.

## 2.4 2004 Census Test

The 2004 Census Test was conducted in portions of northwestern Queens County in New York and in Colquitt, Thomas, and Tift Counties in Georgia. The 2004 Census Test CRFU contained a first attempt at operationalizing the complex problem of unduplicating the census. The CRFU operation used a followup paper questionnaire to investigate person duplication problems, as well as other types of coverage issues. The questions asked of the respondents were similar to those asked in the 2003 LDD study. There was a distinct area on the questionnaire for the enumerators to complete when they encountered housing-level duplication, called the housing unit assessment box.

Since the 2004 test was a site test, the results do not include as many geographically-complex issues of duplication as exist in the decennial census. For instance, site tests do not capture the multiple residences of a college student who goes to school in a different state from where their parents reside or ‘snowbirds’ who spend part of the year in a northern climate and part of the year in a southern climate. There was also no enumeration of persons in GQs during the 2004 Census Test.

The computer-matching algorithm was implemented to match the census universe against itself and identify potentially duplicated persons. Person matches above the identified cutoffs were considered likely duplicates.

Followup work was conducted both with telephone interviewing and with fieldwork. The philosophy for determining which potential duplication cases were sent for a telephone interview and which were sent directly to the field was based on the belief that a telephone interview was sufficient for the resolution of person-level duplication problems, but it was inadequate for housing-level duplication problems. Housing-level duplication can only be resolved by visually seeing the location, justifying the more expensive field-work, while complex living situations associated with person-level duplication can be resolved through a phone conversation.

Since some cases can be resolved with a personal telephone interview and some require fieldwork to be resolved, it was necessary to pre-determine which cases were likely housing-level problems that should be sent to the field. Whole-household duplication was hypothesized to be associated with housing-level problems. Whole-household duplication occurs when a link is made between two households in which *all* of the people in one household match to *all* of the people in the other household. In other words, it was thought that if the duplication was a result of housing-unit duplication, then the whole households would be duplicated as well. In the 2004 CRFU operation, cases of whole-household to whole-household (WW) or of whole-household to

partial-household (WP) duplication were sent directly to the field, while partial-household to partial-household cases (PP) were sent first to telephone resolution.<sup>4</sup>

In 2004, housing unit links that were sent to telephone followup and had the same telephone number were physically clipped together for the interviewers. Linked cases that were sent to field followup and were in the same block were also clipped together. This resulted in increased efficiency for the interviewers and reduced respondent burden, though it still required two interviews for the household – one for each census return.

As documented in Pennington (2005a) and Marshall (2008b), the results from the 2004 Census Test were promising. Notable conclusions were:

- Approximately two-thirds of the duplication cases were determined to be a result of either form misdeliveries or housing unit duplication. These cases could be resolved in an operation designed to determine how many housing units are in existence.
- Some of the person-level cases were difficult to resolve, but improvements in the questionnaire and in the operation could yield greater success, particularly for those cases that are duplicated over longer distances.
- WW matches inside the same census block were most often caused by housing-level duplication issues.
- Results from WP and PP matches did not agree with the research hypothesis but the results did produce new hypotheses to test in the next census test. In particular, the level of household duplication did not prove to be strongly associated with housing-level problems. Geographic distance appeared to be a correlating factor.
- Interviewers had trouble separating the concepts of persons and addresses in such a person-focused field operation. They often began by finding the people instead of finding the address.

Due to the 2004 Census Test results, two separate operations were planned for the 2006 Census Test; one operation would include person-level duplication cases while the second focused on housing-level duplication cases.

## 2.5 2005 National Census Test

The 2005 National Census Test (NCT) was a nationally representative sample of housing units in mailout/mailback areas. The portion of the 2005 NCT that related to coverage improvement efforts contained 210,000 housing units. The chance of finding a person duplicated on these forms was small since both forms would have to fall in the 2005 NCT sample. Thus, an unduplication effort was not conducted since it was neither practical nor promising.

However, a telephone-only CFU<sup>5</sup> operation was still conducted on other types of coverage improvement cases using an automated instrument. Enhancements to the CFU interview and

---

<sup>4</sup> Partial-household to partial-household cases (PP) were matches made between two households where *some* of the people in one household matched to *some* of the people in the other household. A whole-household to partial-household case (WP) was where a match was made between two households so that *all* of the people in one household matched to *some* of the people in the other household.

<sup>5</sup> The operation known as CRFU in 2004 was referred to as CFU in 2005 and the rest of the decade.



operation as a result of the 2005 NCT were considered to be beneficial to the future resolution of all coverage cases, including duplication cases.

## **2.6 2006 Census Test**

The Census Bureau conducted the 2006 Census Test in two sites; Travis County, Texas and the Cheyenne River Indian Reservation and Off Reservation Trust Lands in South Dakota. Building on lessons learned from the 2004 Census Test, there were two followup operations to correct coverage errors: the CFU interview and the Housing Unit Verification (HUV) operation.

The 2006 CFU interview was an improved version of the 2004 CRFU and the 2005 CFU interviews. The interview used an in-depth questionnaire asking about types of living situations to resolve person-level coverage issues, including person-level duplication. The interviews were initially done by telephone, but if the telephone interview was not successful, a field interview was conducted. The telephone interviews used a revised automated instrument from the 2005 NCT, while the field interviews used a revised paper questionnaire.

HUV was a new address list development operation designed to resolve housing-level duplication issues. Resolving issues with the address list in turn helps to resolve the duplicated people. Since HUV was an address list development operation, HUV cases only received a personal visit. HUV enumerators visited the addresses to verify whether the address existed, did not exist, or was a duplicate of another address. This operation was designed to mimic the FV operation that takes place at the end of the decennial census to determine the status of units.

As in the 2004 Census Test, computer matching was used to identify duplicates in the 2006 Census Test. For the first time, the universe of household returns was also matched against the universe of GQ returns to identify people who may have been duplicated between the GQ enumeration and the housing unit enumeration. However, of the 3,409 people who were identified as potential duplicates after all processing was completed, only 11 were counted in both a housing unit and a GQ (Krejsa et al, 2007). Once these person matches were identified, the housing unit responses involved in the match were sent for followup.

Since housing-level duplication issues cannot be resolved through a telephone interview and person-level duplication issues cannot be resolved through an address-listing operation, it was necessary to pre-determine whether to send a case to the CFU interview or to the HUV interview. This determination incorporated the geographic distance between duplicates for the first time. The classification of how much of the household was matched between the two addresses (WW, WP, or PP) was maintained in order to confirm the results seen in 2004.

All within-block WW links and any within-block clusters that involved three or four housing units were eligible for HUV only. Since the 2004 Test found that WP and PP within-block links were more likely to also be housing-level duplication issues, the 2006 unduplication sample selection was designed to send half of those matches for a CFU interview and half to the HUV operation as a way to test the hypothesis that they were housing-level duplication issues. Housing units involved in matches that were not within the same block were eligible only for CFU and not HUV (regardless of whether they were WW, WP, or PP matches).

The results of these efforts in 2006 were both disappointing and encouraging (Krejsa et al, 2007). For cases sent to CFU, the interview was unable to confirm the person as being a duplicate in over ninety percent of cases. Since only a handful of links were found between a housing unit and a GQ, no conclusions could be reached about the utility of the CFU interview to resolve such cases. More encouragingly, the HUV operation confirmed that there was a greater benefit to census coverage by sending the within-block cases to HUV than to CFU.

## 2.7 2008 Census Dress Rehearsal

The 2008 Census Dress Rehearsal was conducted in two sites: San Joaquin County, California and a nine-county area surrounding Fayetteville, North Carolina. The initial plans for the 2008 Census Dress Rehearsal were scaled back in fall 2007 due to budget and schedule constraints, thus impacting the scope of the unduplication program. Notably, the Nonresponse Followup (NRFU) and GQE operations were canceled from the 2008 Census Dress Rehearsal, thus limiting the duplication universe to only mail returns. The 2008 Census Dress Rehearsal was also to include a post-enumeration look at housing units on the address list through the FV operation. However, FV was canceled entirely during the re-design of the 2008 Census Dress Rehearsal; the CFU operation remained in place and all eligible returns containing a duplicated person were sent to CFU.

As with the previous mid-decade census tests, the universe of housing unit returns was matched against itself to identify people who may have been duplicated. Scores were assigned to the quality of a match and cutoffs were established. For person matches above the cutoffs, the responses from which these person matches came were identified and the descriptors of geographic distance between linked responses were altered from previous tests to encompass three categories:

- The housing units were located within the same geographic block.
- The housing units were located within two different blocks, which are adjacent to each other, called surrounding blocks.
- The housing units were “beyond the surrounding block” from each other.<sup>6</sup>

Research from the 2004 and 2006 Census Tests showed that when persons in one household matched to persons in another household within the same geographic block, it was primarily caused by housing-level duplication issues. Therefore, duplication cases that had a within-block match were to be sent to the 2008 FV operation. Since FV was canceled, these cases did not receive any followup interview in the 2008 Census Dress Rehearsal and thus no additional conclusions could be drawn.

The duplication cases with a geographic match from surrounding blocks were initially sampled so that some went to CFU and some to FV. This was done to research which operation was most successful at resolving the duplication and therefore determining which type of duplication issue was the primary issue in surrounding block cases – either housing-level duplication issues or person-level duplication issues. When the FV operation was canceled, the sampling rate was

---

<sup>6</sup> ‘Beyond the surrounding block’ meant that a link was within county (but not a surrounding block link), within state, or within U.S.

changed and all surrounding block cases were sent to CFU. This allowed only for testing the hypothesis that CFU would not resolve a high percent of these cases.

The duplication cases with a geographic match beyond the surrounding block were thought to be primarily caused by person-level duplication issues and therefore were automatically sent to CFU. The 2008 CFU Assessment (Govern, Kostanich, and Heime1 2009) does not report the total number of potential duplicates found from the person matching process; it reports only the number of potential duplicates (2,641) that were in a housing unit that completed a CFU interview.

Results from the Dress Rehearsal confirmed that CFU was particularly unsuccessful at reaching a resolution for surrounding block cases, as the CFU interview was tailored towards issues of person-level duplication (Govern, Kostanich and Heime1 2009). Cases that matched beyond the surrounding block were more successfully resolved during the CFU interview, but the final results showed almost 80 percent of the potential duplicates still were not resolved during CFU (Govern, Kostanich and Heime1 2009). While beyond the surrounding block cases were more likely to be caused by person-level duplication issues, CFU still did not have much success at resolving their residency.

### *Cognitive Testing*

In light of the low success rate from CFU interviews, an experiment was planned for the 2010 Census whereby additional probing questions would be asked after unsuccessful CFU interviews. These questions were cognitively tested in 2008, with one series of probes for duplication cases from the 2008 Census Dress Rehearsal (Childs et al, 2009). The experimental wording of the tested questions differed depending on whether the interview was conducted with the actual suspected duplicated person or another household member. When speaking with the actual suspected duplicate, the tested question wording mentioned the state in which we thought the person might have been duplicated. The wording used when speaking to a proxy did not mention the name of the state in which the person was duplicated, but rather used the generic wording, “another residence.” Since the duplicates were identified through a site test, all duplicates were identified within the same state. Thus, it was not possible to test out-of-state duplicate question wording, except through a hypothetical scenario.

The first question in the experimental series identified the duplicated person by name and said: “NAME may have been counted at another residence (fill: ‘in STATE NAME’) as well as on your census form.” Researchers found this phrasing was very sensitive, especially for one respondent – a father with sole custody of his children – who was upset because he inferred from this question that the mother of his children had reported the children on her census form. In this case, we believe we inadvertently provided information that allowed him to make a connection that may have been perceived as a breach of confidentiality. Based on the interview, we believe that he told the mother of the children about the interview and would have accused her of completing her census form incorrectly. There were two other cases involving duplicated children where a similar conversation may have happened following the interview. In one case, the respondent told the interviewer during a post-interview telephone conversation that she had called her daughter about the situation (the interviewer called the respondent back to ask a

followup question), and in another case, based on the conversation, we believe that the respondent was going to discuss the situation with the duplicated child's parent. Though this sensitive situation was observed with duplicated children, this situation could also happen with adults. In one case of an adult son who sometimes stayed with his girlfriend, the respondent (his mother) made some very critical and hostile remarks about her son's girlfriend and reported inconsistently about how much time the son spent with the girlfriend. This was a sensitive reaction as well, and it could also have led to a conversation by which the girlfriend felt her confidentiality had been breached.

From these cognitive interviews, it was concluded that followup questions should not state that records indicate a specific person may have been duplicated. While this text does not violate the Census Bureau's confidentiality mandates, it could be perceived as violating confidentiality if the respondent figures out who duplicated them. Instead, the recommendation for future experimental questions was to put less focus on people being counted somewhere else and more focus on giving examples of types of places where a person could have been counted that have not been mentioned in CFU previously, such as a parents' house or a girlfriend's or boyfriend's house.

## 2.8 2010 Census

As with the mid-decade census tests, the universe of housing unit returns in the 2010 Census was matched against itself to identify people who may have been duplicated. GQ returns were also included. However, returns were only included if they were data captured by the end of July 2010 and were in scope for CFU. The 2010 CFU universe consisted of responses from the following initial census returns:

- Mailout/Mailback (including Bilingual, replacement mailings, Fulfillment, and Experimental),
- Update/Leave (U/L),
- Enumerator Questionnaires<sup>7</sup>,
- and Telephone Questionnaire Assistance (TQA) interviews.

Scores were assigned to the quality of a match and cutoffs were established. For person matches above the cutoffs, the responses from which these person matches came were identified and the geographic proximity of the matches was established. There are five levels at which links can be geographically associated. The levels below are mutually exclusive and are listed in order of precedence:

- a. Within the same block
- b. Within surrounding blocks of each other
- c. Within the same county
- d. Within the same state
- e. Outside of the state

---

<sup>7</sup> Includes Nonresponse Followup, Nonresponse Followup Reinterview, Nonresponse Followup Vacant Delete Check, Nonresponse Followup Vacant Delete Check Reinterview, Update/Enumerate, Update/Enumerate Reinterview, Remote Alaska, and Remote Update/Enumerate returns

Within-block and within-surrounding-block links are considered to be address-level duplication issues and are thought best resolved in a field operation focused on address resolution, such as FV. Links beyond the surrounding block are considered to be a result of person-level duplication issues and best resolved with an interview focused on living situations, such as CFU.

There was both a FV operation and a CFU operation as part of the 2010 Census, but not all potential duplication cases were sent for followup in those operations due to budget, schedule, and infrastructure constraints. Only certain types of housing-level duplication were researched in FV while a sample of person-level duplication was selected for research in CFU. The CFU results were not used in final Census household determination unless the cases were included for other coverage reasons.

The CFU operation was a vehicle to resolve many different types of coverage issues beyond duplication, including:

- households with more people reported in the household population count box than were able to be fully detailed on the census return<sup>8</sup>,
- returns that had a count discrepancy<sup>9</sup>,
- returns that flagged the undercount or overcount question, and
- returns flagged by Administrative Records processing (more information on each of these cases can be found in Blough 2010).

The undercount and overcount questions (as they appeared on the mailout/mailback questionnaire to housing units) are presented below. The undercount question was for the respondent to indicate, at a household level, that there were additional people staying at the household who were not included in the household population count box. Figure 1 presents this question wording.

**Figure 1: Undercount Coverage Probe**

- 2. Were there any additional people staying here April 1, 2010 that you did not include in Question 1?**  
 Mark ☒ all that apply.
- ☐ Children, such as newborn babies or foster children
  - ☐ Relatives, such as adult children, cousins, or in-laws
  - ☐ Nonrelatives, such as roommates or live-in baby sitters
  - ☐ People staying here temporarily
  - ☐ No additional people

The overcount question was included after the demographic questions on each person panel, for the respondent to indicate if that person sometimes lived or stayed elsewhere. Figure 2 presents the overcount question wording.

<sup>8</sup> These are called “large households.” For instance, this occurs on a mailback form that has more than six people.

<sup>9</sup> Cases where the number of valid people enumerated on the form was different from the provided population count.

**Figure 2: Overcount Coverage Probe****7. Does this person sometimes live or stay somewhere else?**

- ☐ No    ☐ Yes — Mark ☒ all that apply.
- |   |   |
|---|---|
| <input type="checkbox"/> In college housing                   | <input type="checkbox"/> For child custody  |
| <input type="checkbox"/> In the military                      | <input type="checkbox"/> In jail or prison  |
| <input type="checkbox"/> At a seasonal<br>or second residence | <input type="checkbox"/> In a nursing home  |
|   | <input type="checkbox"/> For another reason |

The budget for the 2010 CFU operation funded interview attempts with almost eight million households. However, it was estimated that 22.5 million households representing all possible coverage issues could have been flagged as potentially eligible for followup in the 2010 Census.

A decision had to be made about which coverage issues to send to CFU due to the resource constraints. Count discrepancy cases and large households were included in the 2000 coverage operation and thus considered as the baseline for the 2010 CFU. To determine which additional cases to include in the CFU operation, the Decennial Statistical Studies Division (DSSD) recommended a decision that maximized the expected number of corrections to be made for the money available (Poehler, 2010a and Poehler, 2010b). A “cost per roster change” statistic was calculated from census test results and used to rank the cases. The total number of estimated roster changes for each type of case was calculated as the sum of the number of people to be added to a household roster plus the number of people to be deleted from a household roster as a result of CFU.

To refine the workload for the 2010 Census and allow the possibility of following up with duplicates in either CFU or FV, duplication cases were delineated by geography.

- Within-block and surrounding-block cases were considered eligible for FV and were further delineated based on whether the duplicate link was between two single-unit buildings, two multi-unit buildings, or a single-unit building with a multi-unit building. It was found in our testing that these available variables resulted in different rates of resolution of cases, so only within-block housing unit duplicates that were between a single-unit building and a multi-unit building were sent to FV. This combination of outcomes was thought to be associated with small multi-unit buildings, which can create notoriously difficult enumeration situations. However, there is no variable that identifies small multi-unit buildings.
- Within-county, within-state, and within-U.S. cases were initially considered eligible for CFU. However, the mid-decade tests with within-county duplicates were not successful at resolving duplication so the estimated cost-per-roster-change for those cases was too high to make the cutoff and be included in the 2010 CFU operation. There was no information available from mid-decade research to predict the cost-per-roster-change statistic for within-state and within-U.S. cases, so they also did not make the cutoff for inclusion in the 2010 CFU operation.

Thus, most person links identified as being potential duplicates were not part of a production census operation to improve coverage. Only some of the housing-level cases were included in

FV<sup>10</sup> and a sample of person-level cases were included in CFU for evaluation purposes (please refer to Section 3.3 for more information on the sample design).

The following sources of coverage improvement were included as production case types in the 2010 CFU operation:

- Large Households
- Count Discrepancies:
- Overcount Coverage Probe – “In College Housing” category only
- Overcount Coverage Probe – “In a Nursing Home” category only
- Overcount Coverage Probe – “In Jail or Prison” category only
- Overcount Coverage Probe – “In the Military” category only
- Overcount Coverage Probe – “Household Multiple,” where multiple people on one return marked different overcount categories
- Administrative Records cases - cases where at least one person was matched between an administrative record and the census return for that housing unit and at least one person was identified on the administrative record but not on the census return.
- Undercount Coverage Probe – “People staying here temporarily” category only
- Undercount Coverage Probe – “Relatives, such as adult children, cousins or in-laws” category only
- Undercount Coverage Probe – “Children, such as newborn babies or foster children” category only
- Overcount Coverage Probe – “Person Multiple,” where at least one person on the return marked more than one overcount category
- Undercount Coverage Probe – “Nonrelatives, such as roommates or live-in babysitters” category only

These sources of coverage improvement were called production sources because changes to these cases during the CFU interview could change the makeup of the household in the final 2010 Census count. Alternatively, evaluation cases were case types that mid-decade testing suggested were not as likely to produce roster changes. Cases that only had evaluation case types (such as marking the overcount-seasonal or overcount-custody box) were sampled for the 2010 CFU operation, and the CFU returns of evaluation cases did not affect residency statuses or census counts. For more information on evaluation CFU case types, see Stewart (2010) and Heimerl (2010).

Production cases are also referred to as “above the line” cases since they made the cut to be included in CFU, while evaluation cases are referred to as “below the line” cases.

Information gathered during the initial enumeration was passed to the CFU interview, and respondents were asked about the people rostered on the initial return. The CFU interview contained probes to identify people who were not initially included on the household roster as well as people who, according to the census residence rule, were on the roster but should not have been enumerated at the housing unit. Regardless of the source of coverage improvement,

---

<sup>10</sup> The housing units visited in the FV operation will not be analyzed in this evaluation. For those results please see McPhillips (2012).

all households sent for followup received the same core questions to identify missed and erroneously enumerated people.

### **3. Methodology**

There is much to be learned in the 2010 Census about both person-level and housing-level duplication. However, this evaluation will primarily discuss efforts to understand and resolve person-level duplication. Information about housing-level duplication will be documented only minimally in this evaluation; some additional information was documented in the Field Verification Assessment (McPhillips, 2012).

#### **3.1 Research Components**

In order to gain as much knowledge as possible about person-level duplication and adequately carry out the research objective, three distinct research components were created.

1. Conduct the 2010 CFU interview with a sample of housing units containing potentially duplicated persons.
2. When applicable, utilize experimental questions added to the end of the CFU interview that were intended to learn more about coverage issues unresolved in CFU.
3. Conduct in-depth Qualitative Interviewing (QI) on a variety of duplication cases.

These research components are discussed in more detail below. A fourth component had initially been considered, but did not make it to fruition due to resource constraints. The fourth component would have contacted about 100,000 cases by telephone using the new questionnaire that was developed and ultimately used in the qualitative interviewing.

##### **3.1.1 Conduct the regular 2010 CFU interview with a sample of housing units containing potentially duplicated persons.**

The CFU interview in mid-decade tests was not productive at resolving issues of duplication. One reason could be that the living situations that lead to duplication are casual, infrequent, or loosely defined. Thus, respondents do not believe that the living situation at the root of the duplication warrants a mention to the census interviewer. It is hypothesized that this is especially true for links found within the same county, the predominant type of person-level duplication case that exists in a mid-decade site test. Links that are found farther apart, the long-distance duplicates, had never been tested in the CFU interview. It was hypothesized that they might have a higher rate of resolution since the reason for a duplication to occur over a long distance could have been salient enough for the respondent to mention it to the interviewer.

For instance, college students who attend school close to home might have a dormitory room or apartment near campus but still stay at home often. Such students should get counted at college but might also get counted at their parental home because they are home frequently and the parent still considers them to be a part of the household, thus creating a local duplicate case. The CFU respondent might not mention that the child has another place for college because they do



not consider the college address to be a permanent residence. Meanwhile, college students who attend school in another state from where their parental home is located tend to stay at college full-time during the semester. The child's residence at school is a more salient feature of life for that housing unit so the CFU respondent might be inclined to mention it. Thus, it is hypothesized that long-distance duplicate cases were more likely than local duplicate cases to mention the living situation that produced the duplication.

The mid-decade census tests did not contain any long-distance duplicate links because they were geographically-constrained site tests. To test the hypothesis that long-distance duplicates would perform differently in CFU than local duplicates, a sample of within-state and within-U.S. duplicate links were sent for a 2010 CFU interview. A sample of within-county duplicates were also sent to CFU in order to compare results from the decennial census to the results from the mid-decade site tests. Additionally, a sample of HU-GQ links were sent to CFU, since few HU-GQ cases were able to be identified during the mid-decade tests. The GQ was not contacted for followup in the 2010 CFU; only the housing unit was eligible for a CFU interview.

This research component was included as part of the CFU operation. CFU was conducted from April through August 2010 but due to the timing of the computer-matching process, the interviews with the sampled duplicate cases were conducted from June through August 2010.

### **3.1.2 When applicable, utilize experimental questions at the end of the CFU interview.**

The CFU interview was designed to resolve all person-level coverage issues, not just those related to duplication. For instance, returns could also become eligible for CFU if the respondent marked the undercount question on the initial census enumeration, indicating that a person was potentially left off their census roster. Additionally, returns could become eligible for CFU if the respondent marked the overcount question on the initial census enumeration, indicating that a person sometimes lived or stayed at another place.

The CFU interview was able to resolve a percentage of undercount and overcount cases, thus determining whether a person was indeed missed or possibly counted more than once. However, there were cases where the CFU respondent did not resolve the coverage issue, nor even allude to a complex situation as was indicated on the original census return. For instance, the initial census return could show that a child lived or stayed at another place for a custody situation, but the CFU respondent might never mention a custody situation or even any other possible residence for the child in question.

During the 2010 CFU operation, a series of experimental questions was administered as an exit interview to a sample of respondents who did not mention the complex living situation that we expected them to mention during the standard CFU interview (Childs et al, 2009). This experimental component of the CFU interview, referred to as Mod Q, specifically mentioned what was marked on the original census return (which the 2010 CFU interview did not do) and asked if the respondent remembered what they were thinking about when they answered the original census questionnaire. The answers to these questions were designed to help explain why CFU is not more successful at resolving certain coverage issues (Stewart, 2010).

This evaluation will look at the persons that marked an overcount question and were a duplicated person, then completed a CFU interview and answered the Mod Q questions.<sup>11</sup> The experimental CFU questions should give us insight into why the CFU interview is not always successful at resolving duplication in the census.

This research component was conducted within the CFU operation, conducted from April through August 2010.

### **3.1.3 Conduct in-depth Qualitative Interviewing (QI) on a variety of duplicates.**

The research on duplicated people since Census 2000 has all been acquired from scripted interviews that lead to quantitative results. There is little research available to tell the full story of a duplicated person's living situation (Childs et al, 2009). Thus, a qualitative study was conducted that consisted of 276 cognitive interviews with sampled duplicates. Most of the interviews used a new questionnaire followed by a debriefing; other interviews were semi-scripted and resembled an open-ended dialogue.

The new questionnaire included blended and modified questions from the existing CFU interview as well as from Census Coverage Measurement<sup>12</sup> interviews, in addition to introducing new questions. The goal was to test new questions, which would hopefully prompt the respondent to tell us about the living situation that caused the suspected duplication. The questions were targeted to be relevant to each case, so the new questionnaire was referred to as the Targeted Coverage Followup (TCFU) questionnaire. In the absence of a field test between the 2008 Census Dress Rehearsal and the 2010 Census, these new questions were further developed through cognitive testing with respondents who simulated those who could be duplicated in the census (Childs et al, 2011).

The TCFU questionnaire was different from the CFU interview in a number of ways.

1. For linked cases that had the same telephone number, the TCFU questionnaire asked the respondent to clarify which address they were currently at, before asking further questions about both addresses. In contrast, the CFU interview would make two telephone calls, one for each distinct address.
2. The TCFU questionnaire asked questions specifically about each potential duplicate, whereas the CFU interview asked broadly whether anyone in the household had stayed somewhere else.
3. The TCFU questionnaire was tailored so that different questions were asked of potential duplicates who linked between an HU and a GQ, as compared to potential duplicates who linked between two housing units. HU-GQ duplicates were asked questions specifically about the type of GQ where they were counted, while HU-HU duplicates were asked

---

<sup>11</sup> The Alternative Coverage Followup Questions and Design Evaluation will discuss the general results from the Mod Q questions, but will not analyze duplication cases separately. This evaluation will extend the analysis and discuss the results of duplication cases in Mod Q (Stewart, 2010).

<sup>12</sup> The Census Coverage Measurement Person Interview (CCM PI) and Person Followup (CCM PFU) interviews were referenced in the development of this questionnaire.

questions based on their age (categorized either as a child, an adult, or a senior citizen). For instance, it was asked if children stayed somewhere else for childcare purposes.

4. The questionnaire specifically asked whether potential duplicates stayed with relatives, friends, or even significant others. These relationships were not mentioned as explicitly in the CFU interview.

Prior cognitive testing supported that these changes would lead to more admissions from respondents about other places where they (or the potential duplicate) sometimes lived or stayed (Childs et al, 2011). The debriefing that followed the interview with the TCFU questionnaire explored how respondents reacted to the questions, with particular attention paid to whether respondents felt that we were implicitly revealing confidential information.

Additional semi-structured qualitative interviews were conducted without using the TCFU questionnaire. These interviews collected more holistic information from the respondent about their experience with the decennial census and elicited information on their living situation in a less structured manner.

The Census Bureau's Statistical Research Division (SRD), DSSD, and the Decennial Management Division (DMD) managed contractors hired to conduct the in-person interviews in five metropolitan areas across the country. The potential duplicates who were interviewed represented all ages, included a variety of household sizes, included phone match cases, and included HU-GQ cases. The research explored the different types of duplicate situations, how respondents thought about their living situations and what sensitivities existed, thus ultimately providing more information on why the current CFU interview is not successful at resolving person-level duplication.

This QI research began in fall 2010 and finished in September 2011. Excerpts of the full report are included in this evaluation; for the complete report, refer to the report by Peytcheva and colleagues (2011).

## **3.2 Research Questions**

The following high-level questions describe the information and results that are included in this evaluation. The first section contains information on the universe of cases identified as being potential duplicates, followed by results of each of the three afore-mentioned research components.

### **3.2.1 Universe of Duplication Cases**

What did the universe of duplication cases look like?

- a. How many potential duplicate matches were identified during processing?
- b. What was the eligible, sampled, and completed workload for each component of this evaluation (Production CFU, Mod Q, QI), including the cases that were eligible for FV?

- c. What was the demographic profile of people and returns identified as potential duplicates?

### **3.2.2 CFU Sample**

How successful were the cases sent to Production CFU?

- a. How many duplicates acknowledged having a complex living situation?
- b. How many duplicates were either deleted or linked to a deleted person?

### **3.2.3 Experimental CFU Questions (Mod Q Cases)**

What information was obtained about the living situation of duplicates from the cases that went into Mod Q?

### **3.2.4 Qualitative Interviews**

What results were obtained from the cases that completed a Qualitative Interview?<sup>13</sup>

The research questions will be analyzed and presented in a variety of cross-tabs, including by geography of match, type of duplicate case (HU-HU or HU-GQ), and demographic characteristics, among others.

## **3.3 Sample Design**

The method in which the sample was selected for each research component of the Effectiveness of Unduplication Evaluation was defined by DSSD (Marshall, 2009b).

At a high level, the sample for each research component was based on the universe of person links identified using the computer-matching algorithm. Before the sample could be selected for each component of this evaluation, the universe had to be prepared. The following steps were taken to prepare the duplication universe:

1. Using the person links identified as a result of computer matching, response links were identified.
2. Each response link was associated with a geography that describes the distance between the two responses. One of the following five geographies was identified for each link:
  - a. Within block response link
  - b. Surrounding block response link
  - c. Within county response link

---

<sup>13</sup> This research was thoroughly analyzed by the contractors hired for this component. Excerpted results are presented in this report.

- d. Within state response link
- e. Outside of state response link
- 3. For each response link identified as either a within county, within state, or outside of state response link, the following was determined:
  - a. Whether the response link represented an HU-HU response link or an HU-GQ response link.
  - b. Whether the response link represented a coverage experiment response link. This meant that at least one of the HU forms in the response link was from the D-1 (X13) form, an experimental mail questionnaire that had an expanded series of overcount questions.<sup>14</sup>
  - c. Whether the response link represented a GQ experiment response link. This meant that the GQ form in an HU-GQ response link was from the D-20(X1) form, an experimental GQ questionnaire that attempted to collect a secondary address for each respondent.<sup>15</sup>
  - d. Whether the response link represented an overcount overlap response link. This meant that at least one person record on one of the HU responses affirmatively marked the overcount question.<sup>16</sup>
  - e. Whether the response link represented a duplication-only response link. This meant that neither half of the response link had a positive mark for the overcount question or was linked to either the D-1 (X13) form or the D-20(X1).
  - f. Whether either response in the response link was associated with a Master Address File ID (MAFID) that was sent to production CFU in an earlier wave. More precisely, whether
    - i. both responses were associated with a MAFID that was sent to CFU in an earlier wave,
    - ii. only one response was so associated, or
    - iii. neither response was so associated.

Ultimately, only response links within county, within state, or outside of state were eligible for this evaluation's sample selection, out of the five possible levels of geographic match (identified in Step 2 above). Prior to sample selection, response links were removed where both responses were associated with MAFIDs that were sent to CFU in an earlier wave or response links where either of the responses was associated with a Puerto Rico address.<sup>17</sup> Response links were then sampled and stratified based on flags detailed in parts 3a-3e above. The links were sent to CFU if they passed additional eligibility checks.

The remaining within county, within state, or outside of state response links which were duplication cases only and were not sampled for CFU were eligible for the Qualitative Interview

---

<sup>14</sup> For more information on this experimental questionnaire, refer to the Avoid Followup study plan (Jackson and Heibel, 2010).

<sup>15</sup> For more information on this experimental questionnaire, refer to the Avoid Followup study plan (Jackson and Heibel, 2010).

<sup>16</sup> More defined flags are set based on which overcount category is marked. Details are provided in Marshall (2009b).

<sup>17</sup> Puerto Rico addresses were removed from sample selection because the experimental interviews were conducted in English only.

component. Not all of these response links were contacted, depending on the metropolitan areas chosen and the type of duplication cases needed for the study.

### **3.4 Sample Weighting**

Weights were used in the analysis of CFU results. The weights for each sampled link were created assuming simple stratified random sampling.

### **3.5 Data Sources**

The data for this evaluation came from multiple files which are described below.

#### **3.5.1 Duplicate Person Identification (DPI) Files**

These files were produced by DSSD at both the housing unit and person levels. They identified all response links and person links from the matching process that were above the identified cutoff.

#### **3.5.2 2010 Decennial Response File (DRF)**

The DRF contained the core response data that made up the Universal Response Database from all questionnaires that were data captured in the 2010 Census. This file provided information on demographic characteristics of the person and response links. Decennial Systems & Processing Office created the DRF.

#### **3.5.3 Unduplication Analysis Files**

The unduplication analysis files merged DRF data onto the DPI files as appropriate.

#### **3.5.4 CFU Assessment Files**

This file provided information on the results of the CFU operation. It also included housing unit and person data from the DRF (and from auxiliary data) for all cases selected and sent to the CFU operation. The CFU Assessment Files were created by DSSD.

#### **3.5.5 Address File**

This file was a subset of the Universe Control and Management table pulled during 2010 Census operations and contained the ZIP code information necessary for select tables in this analysis.

## **4. Limitations**

There are a number of limitations to this evaluation.

### **4.1 Incomplete Universe of Potential Duplicates**

The DPI files were compiled using returns that were data captured by late July 2010. Some census returns were not data captured at that point and were not available to be included in the matching algorithm. This included a number of GQ returns as well as cases enumerated in the final field operations.

The matching algorithm did not include any census returns from the non-ID universe. For instance, Be Counted Forms, Enumeration at Transitory Locations forms, added housing units from field operations, and some interviews conducted in the Telephone Questionnaire Assistance operation were excluded from the matching. Additionally, any enumerator form that would have required linking to a continuation form (i.e., having more than five people enumerated at the housing unit) was not included in the matching.

Since these returns were not part of the computer-matching algorithm, the universe of potential duplicates should not be taken as a complete picture of duplication in the 2010 Census. Additionally, no GQ-GQ duplication was identified for this research.

### **4.2 Mod Q Cases**

Cases were sampled for Mod Q based on the overcount and undercount probes. The sample was not created with duplication cases in mind, so this universe and results might not be representative of all duplication cases. In addition, the coding results of the responses to open-ended questions were not available for inclusion in this report.

### **4.3 Large Clusters**

The sampling done for CFU only identified links between two returns, and the data files were structured to identify pairs of links. Analysis on links involving three or more responses (large clusters) was not able to be completed for this evaluation.

### **4.4 Recall Bias**

The CFU interview was conducted weeks or even months after a respondent completed the census form. The qualitative interview was conducted months, possibly a year, after the respondent completed the census form. This lag between the initial census response and the followup interview could have impacted respondents' ability to recall information on what they and their household were doing in spring 2010.

## 4.5 Geographic Descriptors

Categorizing geographic proximity of links (within county, within state, etc.) into either a local or a long-distance duplication is an imperfect classification. For example, a person counted in Washington, DC, and in Arlington, VA, is considered a long-distance duplicate because the two responses are located in different states, though the exact addresses could be within a mile of each other.

## 5. Results

The following section answers the research questions posed in Section 3.2. The results are presented in the order that the research questions were presented in Section 3.2, starting with a discussion of characteristics of the universe of potential duplicates, followed by results from the CFU interview and the Mod Q interview, and concluding with the results of the qualitative interviews.

### 5.1 Characteristics of the Unduplication Universe

This section describes the potential universe of cases that was identified by DPI for possible use in the CFU operation and qualitative interviews. As mentioned in Section 4, some census returns were out of scope or were not data captured in time to be included in the matching, so this universe does not represent all the duplicates that might be identified if every census return was eligible for matching. The purpose of this section is to show what characteristics are seen among suspected duplicates. Thus, this section does not identify whether the duplication was eventually resolved in post-processing; either through the CFU interview, through GQ Usual Home Elsewhere (UHE) matching<sup>18</sup>, or whether the Primary Selection Algorithm (PSA)<sup>19</sup> selected another return for the MAFID which did not contain the duplication.

---

<sup>18</sup> As described in the residence rules (Appendix B), persons enumerated in some specific types of GQs were allowed to claim a UHE. The GQ questionnaire had space for the person to provide the address where they lived or stayed most of the time. If the address was able to be geocoded and matched to a MAFID, then HQ processing moved the person from the GQ into the housing unit return (if they had not already been counted there).

<sup>19</sup> When more than one return was data captured for a housing unit MAFID, then the PSA was executed to select a single return to represent that housing unit.



### 5.1.1 Overview

In order to define the universe of potential duplicates for this evaluation, it is important to know that a single MAFID (either a housing unit or GQ) could have had more than one census return. For instance, a housing unit might have mailed in their census questionnaire the day before a nonresponse enumerator knocked on their door and insisted on completing an interview. Both questionnaires would then have been received by census, data captured, and added to the DRF. These returns could have enumerated the same persons or different persons. Additionally, during the census data capture process, the Decennial Response Integration System/Headquarters (HQ) identified that a number of returns had problems and were subsequently “reprocessed” to fix the errors. These “reprocessed” returns were added to the DRF as unique returns, but were largely the same as the original return. The DPI process did not match across returns within the same MAFID, so it is not possible to identify when multiple returns at the same MAFID are in fact the same person since this was not in scope for the DPI process.

We did not want to include links in our universe for this evaluation that represented the same MAFIDs, since they were likely reprocessed or redundant returns. Since any reprocessed return would be the preferred return in census processing, we kept any links with a reprocessed return and discarded links involving the non-reworked returns. Any multiple links that remained between the two distinct MAFIDs were sorted, and one of the links was chosen to represent that MAFID pair for analysis.<sup>20</sup>

A total of 7,784,286 person links were identified by the DPI process. After removing multiple links at the same MAFIDs as described above, Table 1 shows that a total of 7,454,171 person links remained on our dataset for evaluative purposes<sup>21</sup>. The majority of these links (88.5 percent) were identified between two housing units. Two housing units could be linked together with more than one person duplicated between their returns. Each GQ return only enumerated a single person, so by definition there could only be one person from a housing unit response who linked to a distinct GQ response. A total of 4,907,707 response links were identified by the DPI process, but 4,711,560 links of living quarters remained after removing multiple links at the same MAFID. Table 1 shows that 81.9 percent of links at the response level were between two housing units.

<b>Table 1. Number of Response and Person Links by Link Type</b>				
<b>Type of link</b>	<b>Response Level</b>		<b>Person Level</b>	
	<b>Number</b>	<b>Percent</b>	<b>Number</b>	<b>Percent</b>
HU-HU links	3,857,604	81.9	6,600,215	88.5
HU-GQ links	853,956	18.1	853,956	11.5
<b>Total</b>	<b>4,711,560</b>	<b>100.0</b>	<b>7,454,171</b>	<b>100.0</b>

Source: Unduplication Analysis Files

<sup>20</sup> This research was done using data without regard to operations such as CFU or FV or post-processing decisions such as PSA or GQ UHE matching, which may have independently resolved the duplication.

<sup>21</sup> Given the limitations and timing of this research, this figure is not meant to be a final assessment of the amount of duplication seen in the 2010 Census.

Person links can be analyzed either at the response level or at the person level, as shown in Table 1. The magnitude of person-level duplication describes the extent of duplication in the census. However, the magnitude of links at the response level describes the entire workload for any followup operation. This evaluation will present some tables at the response level and other tables at the person level.

### 5.1.2 Geographic Distance

The geographic distance between links is used to identify whether the likely cause of duplication is address-related or person-related. The distribution of geographic proximity is shown in Table 2.

**Table 2: Geographic Relationship of Links**

<b>Geography of link<sup>22</sup></b>	<b>Response level</b>		<b>Person level</b>	
	<b>Number</b>	<b>Percent</b>	<b>Number</b>	<b>Percent</b>
Within block	1,200,553	25.5	2,495,776	33.5
Surrounding block	373,167	7.9	786,273	10.5
Within county	1,480,767	31.4	2,085,442	28.0
Within state	1,061,878	22.5	1,304,804	17.5
Across state	595,195	12.6	781,876	10.5
<b>Total</b>	<b>4,711,560</b>	<b>100.0*</b>	<b>7,454,171</b>	<b>100.0</b>

Source: Unduplication Analysis Files

\*May not sum to 100.0 due to rounding.

Table 2 shows that when two responses are linked, 25.5 percent of the links are within the same block and 31.4 percent of the links are within the same county. When examining the percent of persons linked, 33.5 percent of person links are within the same block and 28.0 percent are within the same county. Thus, even though a smaller percent of responses were linked within block than within county, more persons were identified as linked within block. This can be attributed to the fact that response links within block are more likely to have multiple persons involved in the link, whereas within county they are more likely to involve only one person (see Table 10). It may also be attributable to the fact that more whole household links are found within block (see Table 12).

Table 3 shows the distribution of geographic proximity of person links by the type of link.

<sup>22</sup> The mutually exclusive geographies that are used in the matching process are within block, within tract, within county, within state, and across state. Surrounding blocks are identified after the matching process. Thus, any cases identified as within-tract by matching are redistributed into either surrounding block or within-county categories here.

**Table 3. Distribution of Person Links by Geography and Link Type**

<b>Geography</b>	<b>Total</b>		<b>HU to HU</b>		<b>HU to GQ</b>	
	<b>Number</b>	<b>Percent</b>	<b>Number</b>	<b>Percent</b>	<b>Number</b>	<b>Percent</b>
Within Block	2,495,776	33.5	2,402,030	36.4	93,746	11.0
Surrounding Block	786,273	10.5	773,632	11.7	12,641	1.5
Within County	2,085,442	28.0	1,841,649	27.9	243,793	28.5
Within State	1,304,804	17.5	942,854	14.3	361,950	42.4
Across State	781,876	10.5	640,050	9.7	141,826	16.6
<b>Total</b>	<b>7,454,171</b>	<b>100.0</b>	<b>6,600,215</b>	<b>100.0</b>	<b>853,956</b>	<b>100.0</b>

Source: Unduplication Analysis Files

For person links between two housing units, 36.4 percent of person links were located within the same block and 27.9 percent of person links were within the same county. This is similar to the overall percentages. Links found within-block are suspected to be duplicate addresses on the MAF or potential form misdeliveries, as described in Section 2.1. The HU to GQ person links show a different trend, with 42.4 percent of all links showing up within the same state (but not in the same county). The fact that 11.0 percent of HU to GQ person links occurred within the same block may be an indication that an address was represented with both a HU record and a GQ record, and so was duplicated.

Instead of person links, Table 4 shows the distribution of geographic proximity by the type of link for response links.

**Table 4. Distribution of Response Links by Geography and Link Type**

<b>Geography</b>	<b>Overall</b>		<b>HU to HU</b>		<b>HU to GQ</b>	
	<b>Number</b>	<b>Percent</b>	<b>Number</b>	<b>Percent</b>	<b>Number</b>	<b>Percent</b>
Within Block	1,200,553	25.5	1,106,807	28.7	93,746	11.0
Surrounding Block	373,167	7.9	360,526	9.3	12,641	1.5
Within County	1,480,767	31.4	1,236,974	32.1	243,793	28.5
Within State	1,061,878	22.5	699,928	18.1	361,950	42.4
Across State	595,195	12.6	453,369	11.8	141,826	16.6
<b>Total</b>	<b>4,711,560</b>	<b>100.0</b>	<b>3,857,604</b>	<b>100.0</b>	<b>853,956</b>	<b>100.0</b>

Source: Unduplication Analysis Files

Similar trends exist in Table 4 as were seen in Table 3. The biggest difference is that at the HU to HU response link level, more links occurred within county (32.1 percent) than within block (28.7 percent). This can be attributed to the fact that response links within block are more likely to have multiple persons involved in the link, whereas within county they are more likely to involve only one person (see Table 10).

As an investigative measure, we attempted to use the ZIP Code associated with each MAFID to calculate the distance between them. The exact distance between two addresses was not possible to calculate, but using ZIP code as a proxy, the distance in miles between the center point of each ZIP code was calculated. Thus, addresses in the same ZIP code have a distance of zero. An

“Unknown” distance indicates that at least one of the ZIP Codes associated with the response link did not exist on the data file input, so a distance could not be calculated. This provides an additional gauge of distance between the two linked addresses. A link could be classified as being across state lines, yet one address might be in Manhattan while the other is just across the Hudson River in Hoboken, NJ. This method of describing the distance between the two responses is affected by population density, as regions covered by a single ZIP code are smaller in urban areas than in rural areas. Table 5 shows the distance in miles between each pair of living quarters identified as a duplicate.

<b>Table 5: Distance Between ZIP Codes of Response Links</b>		
<b>Miles Between ZIP Code Centers</b>	<b>Number</b>	<b>Percent</b>
0	1,845,686	39.2
1 to 5	461,963	9.8
6 to 10	422,181	9.0
11 to 100	1,005,791	21.3
101 to 1,000	561,433	11.9
Greater than 1,000	171,630	3.6
Unknown	242,876	5.2
<b>Total</b>	<b>4,711,560</b>	<b>100.0</b>

Source: Unduplication Analysis Files

Table 5 shows that 39.2 percent of response links were in the same zip code. Only 3.6 percent of all response links were more than 1,000 miles apart.

Table 6 shows the distribution of distance for each of the five geographic categories.

**Table 6. Distance Between ZIP Codes of Response Links by Geography**

Miles between ZIP code centers	Overall		Within Block		Surrounding Block		Within County		Within State		Across State	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent
0	1,845,686	39.2	1,078,700	89.9	294,300	78.9	461,264	31.2	11,307	1.1	115	0.0
1 to 5	461,963	9.8	13,614	1.1	19,104	5.1	398,463	26.9	27,702	2.6	3,080	0.5
6 to 10	422,181	9.0	9,428	0.8	11,086	3.0	324,536	21.9	70,910	6.7	6,221	1.0
11 to 100	1,005,791	21.3	9,404	0.8	8,852	2.4	254,651	17.2	644,670	60.7	88,214	14.8
101 to 1,000	561,433	11.9	1,579	0.1	1,213	0.3	547	0.0	257,812	24.3	300,282	50.5
Greater than 1,000	171,630	3.6	100	0.0	111	0.0	31	0.0	59	0.0	171,329	28.8
Unknown	242,876	5.2	87,728	7.3	38,501	10.3	41,275	2.8	49,418	4.7	25,954	4.4
<b>Total</b>	<b>4,711,560</b>	<b>100.0</b>	<b>1,200,553</b>	<b>100.0</b>	<b>373,167</b>	<b>100.0</b>	<b>1,480,767</b>	<b>100.0</b>	<b>1,061,878</b>	<b>100.0</b>	<b>595,195</b>	<b>100.0</b>

Source: Unduplication Analysis Files

Table 6 shows that the distance for links within the same block and for links within surrounding blocks are generally zero miles apart, as expected. Links within the same county are distributed, though the most common distance is also zero miles apart (31.2 percent). Links within the same state are most often 11 to 100 miles apart, while links found across state lines are most often 101 to 1,000 miles apart. Links within the same block and within surrounding blocks had the highest rates of unavailable ZIP codes, perhaps also indicating issues with the addresses that led to the duplication.

### 5.1.3 Form Type

During the 2010 Census, the majority of housing units received a census questionnaire in the mail. Housing units that did not return the initial census questionnaire were then visited by an enumerator. Other housing units never received a questionnaire in the mail by design and were visited by an enumerator. In Table 7 below, we have grouped links into one of three categories.

1. MB – Includes all Mailout/Mailback returns, Update/Leave returns, experimental Mailout/Mailback returns, Fulfillment returns, and Telephone Questionnaire Assistance (TQA) returns.<sup>23</sup>

<sup>23</sup> The number of links with a TQA or Fulfillment return was small so they are included in with MB (all are considered to be self-response modes).

2. ENUM – Includes all enumerator form types, from operations such as NRFU, Update/Enumerate (UE), and Vacant/Delete Check (VDC).
3. GQ – Includes any GQ form types.

**Table 7. Form Types Linked at the Person Level**

<b>Linked Form Type</b>	<b>Number</b>	<b>Percent</b>
MB to ENUM	4,061,825	54.5
MB to MB	1,998,587	26.8
ENUM to ENUM	539,803	7.2
MB to GQ	677,683	9.1
ENUM to GQ	176,273	2.4
<b>Total</b>	<b>7,454,171</b>	<b>100.0</b>

Source: Unduplication Analysis Files

*Note: All links are across unique MAFIDs.*

Table 7 shows that 54.5 percent of all person links occurred between one MB and one ENUM form (it is perhaps useful to reiterate that these links are persons found between two distinct MAFIDs, using different modes to be enumerated, rather than the same MAFID enumerated two distinct times). The final DPI process was executed before some unknown number of enumerator returns had been data captured, which might have increased the proportions duplicated on those forms, especially the proportion between two enumerator-administered returns. The majority of GQ links occurred between a mail form and a GQ form. It is not directly shown in Table 7, but it can be calculated that 79.4<sup>24</sup> percent of all links that involved a GQ were linked to a MB form instead of an ENUM form.

Table 8 shows the distribution by geography for the possible linked form types.

<sup>24</sup>  $(677,683) / (677,683 + 176,273) = 79.4$  percent

**Table 8. Form Types Linked at the Person Level, by Geography**

<b>Geography</b>	<b>MB to ENUM</b>		<b>MB to MB</b>		<b>ENUM to ENUM</b>		<b>MB to GQ</b>		<b>ENUM to GQ</b>	
	<b>Number</b>	<b>Percent</b>	<b>Number</b>	<b>Percent</b>	<b>Number</b>	<b>Percent</b>	<b>Number</b>	<b>Percent</b>	<b>Number</b>	<b>Percent</b>
Within Block	1,922,996	47.3	241,629	12.1	237,405	44.0	52,887	7.8	40,859	23.2
Surrounding Block	582,465	14.3	84,415	4.2	106,752	19.8	8,050	1.2	4,591	2.6
Within County	958,270	23.6	753,558	37.7	129,821	24.0	188,356	27.8	55,437	31.4
Within State	353,462	8.7	543,984	27.2	45,408	8.4	305,513	45.1	56,437	32.0
Across State	244,632	6.0	375,001	18.8	20,417	3.8	122,877	18.1	18,949	10.7
<b>Total</b>	<b>4,061,825</b>	<b>100.0*</b>	<b>1,998,587</b>	<b>100.0</b>	<b>539,803</b>	<b>100.0</b>	<b>677,683</b>	<b>100.0</b>	<b>176,273</b>	<b>100.0*</b>

Source: Unduplication Analysis Files

\*May not sum to 100.0 due to rounding.

Table 8 shows that MB to ENUM person links happened most often within block (47.3 percent of all MB to ENUM links). This could have happened if an address was duplicated on the MAF with minor differences. Two questionnaires would have been mailed out and the respondent completed only one. For example, the second questionnaire was identified for nonresponse followup and an enumerator eventually completed the form for the “other” address.

MB to MB person links happened most often within the county at 37.7 percent. ENUM to ENUM person links happened most often within block at 44.0 percent. MB to GQ person links happened most often within state at 45.1 percent. Lastly, ENUM to GQ person links happened within county and within state at about the same rate, at 31.4 percent and 32.0 percent, respectively.

### 5.1.4 Number of Person Links

Each response link will have at least one person linked between the two responses. For HU to HU links, it is possible to have multiple persons linked between the returns; however, for HU to GQ links there will only ever be a one-to-one link. For HU to GQ links, the decision to only identify one-to-one links was a processing decision since each GQ questionnaire can only enumerate one person and was treated as a unique return, but it is possible to have two college-aged children at the same college link to the same HU. Since all HU to GQ links represent one person, the following table only shows the number of links identified between HU to HU links. The results in Table 9 are presented at the response link level.

**Table 9: Number of Person Links (Per Household) Between HU-HU links**

<b>Number of Person Links</b>	<b>Number</b>	<b>Percent</b>
1	2,257,732	58.5
2	909,464	23.6
3	355,574	9.2
4	228,042	5.9
5	101,066	2.6
Over 5	5,726	0.1
<b>Total</b>	<b>3,857,604</b>	<b>100.0*</b>

Source: Unduplication Analysis Files

\*Does not sum to 100.0 due to rounding.

In over half of the HU to HU response links (58.5 percent), only one person was duplicated. Two people were duplicated on 23.6 percent of all HU to HU response links. Table 10 shows the distribution of the number of person links per HU to HU response link at each of the five levels of geographic proximity.



**Table 10. Number of Person Links Between HU-HU Links, By Geographic Proximity**

Geography	Number of Person Links											
	One		Two		Three		Four		Five		More than Five	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent
Within Block	426,382	18.9	326,027	35.8	165,027	46.4	125,372	55.0	60,704	60.1	3,295	57.5
Surrounding Block	139,028	6.2	107,497	11.8	55,389	15.6	40,376	17.7	17,780	17.6	456	8.0
Within County	863,174	38.2	221,364	24.3	92,861	26.1	42,262	18.5	16,317	16.1	996	17.4
Within State	527,704	23.4	125,280	13.8	28,747	8.1	13,436	5.9	4,192	4.1	569	9.9
Across State	301,444	13.4	129,296	14.2	13,550	3.8	6,596	2.9	2,073	2.1	410	7.2
<b>Total</b>	<b>2,257,732</b>	<b>100.0*</b>	<b>909,464</b>	<b>100.0*</b>	<b>355,574</b>	<b>100.0</b>	<b>228,042</b>	<b>100.0</b>	<b>101,066</b>	<b>100.0</b>	<b>5,726</b>	<b>100.0</b>

Source: Unduplication Analysis Files

\*May not sum to 100.0 due to rounding.

The results show that single person HU-HU links occurred most often within-county or within-state (38.2 percent and 23.4 percent, respectively). When more people were duplicated across the same two HU returns, it was increasingly likely that the links occurred within block.

### 5.1.5 Whole and Partial Households

Another way to define response links is by their household link status. This has historically been referred to as “whole household” or “partial household” status (see Section 2.4 for background), but that description was revised and expanded upon for this evaluation. Each response link in this evaluation is classified as one of the following:

- Single Person (HH): Both linked returns were from housing units, only one person was identified as a CFU Valid person<sup>25</sup> on each return, and that person was identified as a duplicated person. This is essentially a subset of Whole HH matches.
- Whole Household (HH): Both linked returns were from housing units, there was more than one CFU Valid person enumerated on each return, the number of CFU Valid persons was the same on each return, and all of the CFU Valid persons were identified as duplicates.

<sup>25</sup> A person was identified as CFU Valid during Data Capture Audit and Resolution (DCAR) processing. To be considered CFU Valid, a person could not have been identified as a duplicate of another person on the return, and must have provided a valid name or age. For more information on the DCAR process, please see Barrett (2010).

- Partial Household (HH): Both linked returns were from housing units, and the number of person links was less than the number of CFU Valid persons on at least one of the returns.
- Discrepant: Both linked returns were from housing units, but the number of person links was greater than the number of CFU Valid persons on at least one of the returns. This could have happened when there were within-return duplicates across both returns. For example, if a single person received two mailback forms at different addresses, and put their information down for each of the 12 person panels on each form then we would show 12 person links. However, processing would recognize these as within-return duplicates, and thus only one person would be considered CFU Valid.
- Single Person HU-GQ: For housing unit to GQ links, the one person enumerated on a GQ return linked to a housing unit with one CFU Valid Person.
- Partial Household HU-GQ: For housing unit to GQ links, the one person enumerated on a GQ return linked to a housing unit with more than one CFU Valid Person.

Table 11 shows the frequency that each of these situations occurred in the 2010 Census DPI universe.

<b>Table 11: Whole/Partial Status by Type of Link for Response Links</b>		
<b>Whole/Partial Description</b>	<b>Number</b>	<b>Percent</b>
HU to HU links	<b>3,857,604</b>	<b>100.0</b>
Single Person	412,864	10.7
Whole HH (> 1)	1,025,139	26.6
Partial HH	2,417,598	62.7
Discrepant	2,003	0.1
HU to GQ links	<b>853,956</b>	<b>100.0</b>
Single Person	126,224	14.8
Partial HH	727,732	85.2
<b>Total</b>	<b>4,711,560</b>	<b>100.0</b>

Source: Unduplication Analysis Files

Table 11 shows that 62.7 percent of HU to HU links involved partial links. This indicates that only a subset of the persons at the HUs have complex living situations, such as children in joint custody. For HU to GQ links, 85.2 percent involved HUs where there were more persons enumerated than just the person at the GQ.

Table 12 shows the frequency that each of these situations occurred by geographic proximity.

**Table 12. Geographic Distribution of Whole/Partial Status by Response Links**

<b>Geography</b>	<b>Whole HH</b>		<b>Single Person HH</b>		<b>Partial HH</b>		<b>Discrepant HH</b>		<b>Single Person GQ</b>		<b>Partial Person GQ</b>	
	<b>Number</b>	<b>Percent</b>	<b>Number</b>	<b>Percent</b>	<b>Number</b>	<b>Percent</b>	<b>Number</b>	<b>Percent</b>	<b>Number</b>	<b>Percent</b>	<b>Number</b>	<b>Percent</b>
Within Block	535,324	52.2	232,408	56.3	52,313	41.4	1,368	68.3	52,313	41.4	41,433	5.7
Surrounding Block	182,482	17.8	72,359	17.5	4,672	3.7	248	12.4	4,672	3.7	7,969	1.1
Within County	114,217	11.1	49,288	11.9	52,257	41.4	207	10.3	52,257	41.4	191,536	26.3
Within State	84,952	8.3	29,253	7.1	13,960	11.1	106	5.3	13,960	11.1	347,990	47.8
Across State	108,164	10.6	29,556	7.2	3,022	2.4	74	3.7	3,022	2.4	138,804	19.1
<b>Total</b>	<b>1,025,139</b>	<b>100.0</b>	<b>412,864</b>	<b>100.0</b>	<b>126,224</b>	<b>100.0</b>	<b>2,003</b>	<b>100.0</b>	<b>126,224</b>	<b>100.0</b>	<b>727,732</b>	<b>100.0</b>

Source: Unduplication Analysis Files

\* May not sum to 100.0 due to rounding.

Table 12 shows that whole HH and single person HH links occurred most often within block at 52.2 percent and 56.3 percent, respectively, which again is likely to be the result of duplicate addresses on the MAF or form misdeliveries. Partial HH links occurred most often within the county at 44.4 percent, which could have been the result of living situations such as shared custody situations, moves, or tenuously attached persons. Discrepant cases appear to show traits similar to those of whole HH and single person HH links, since 68.3 percent of discrepant cases show up within-block.

Single person GQ links occurred most often within block and within county, both at 41.4 percent. As mentioned before, HU to GQ links found within the same block may indicate that the living quarters is included on the MAF with both a HU and GQ MAFID. Partial person GQ links occurred most often within state at 47.8 percent of all such cases.

### 5.1.6 Phone Match

Every census questionnaire for housing units asked for the respondent's phone number. The increased use of cell phones has been helpful for matching purposes since the phone number listed on linked returns might be the same, even if the respondent was actually at two different addresses. For instance, a household may move to a new address and keep their cell phone as the phone number of record, or a household might have completed a census return for both their primary and secondary home, and listed the same cell phone number on both returns. If a questionnaire was completed by an enumerator using a proxy respondent for an address, then the phone number would have been that of the proxy respondent and not associated with the household members, so those phone numbers

were not included in the matching. Thus, it was highly likely that if two forms had matching telephone numbers, then the forms contained the same persons, even if there was some discrepancy in the person data provided on the form. Phone numbers were validated before persons were sent to the DPI process (Lynch, 2009a).

The following tables indicate how often phone numbers matched between two forms at the response level. Since the GQ forms did not collect a telephone number, the following tables only include HU to HU response links. Blank phone numbers occurred either if a respondent refused to provide a phone number, the phone number was found to be invalid, or the respondent was a proxy.

**Table 13: Frequency of Phone Numbers Matching (HU-HU)**

<b>Phone Number Agreement</b>	<b>Number</b>	<b>Percent</b>
Phone Numbers Matched	949,549	24.6
Phone Numbers Different	2,074,361	53.8
One Number Blank	732,810	19.0
Both Numbers Blank	100,884	2.6
<b>Total</b>	<b>3,857,604</b>	<b>100.0</b>

Source: Unduplication Analysis Files

Over half of the HU to HU links (53.8 percent) had valid phone numbers on both sides but the numbers were different. About one quarter of the links (24.6 percent) had the same phone number however. An additional 19.0 percent of the links only had a valid phone number on one side. Table 14 shows how often phone numbers matched, by the geographic proximity of the links.

**Table 14. Phone Matches at the Housing Link Level by Geography**

<b>Geography</b>	<b>Phone Numbers Matched</b>		<b>Phone Numbers Different</b>		<b>One Number Blank</b>		<b>Both Numbers Blank</b>	
	<b>Number</b>	<b>Percent</b>	<b>Number</b>	<b>Percent</b>	<b>Number</b>	<b>Percent</b>	<b>Number</b>	<b>Percent</b>
Within Block	504,983	53.2	282,782	13.6	266,357	36.3	52,685	52.2
Surrounding Block	191,090	20.1	106,555	5.1	53,396	7.3	9,485	9.4
Within County	123,448	13.0	872,339	42.1	221,974	30.3	19,213	19.0
Within State	72,021	7.6	497,264	24.0	118,847	16.2	11,796	11.7
Across State	58,007	6.1	315,421	15.2	72,236	9.9	7,705	7.6
<b>Total</b>	<b>949,549</b>	<b>100.0</b>	<b>2,074,361</b>	<b>100.0</b>	<b>732,810</b>	<b>100.0</b>	<b>100,884</b>	<b>100.0</b>

Source: Unduplication Analysis Files

Table 14 shows that when the telephone number matched across two linked forms, 53.2 percent of the links were located within the same block. Again, this likely reflects duplication that is caused by form misdeliveries or address duplication. Cooperative respondents would provide the same information, even across multiple enumerations. In links where the telephone numbers did not match, the most common geographic classification was within county (42.1 percent). Table 14 shows that at least one side did not provide a valid telephone number 36.3 percent of the time within block and 30.3 percent of the time within county. Lastly, both sides had blank phone numbers 52.2 percent of the time within block.

Table 15 shows the form type pairings with phone number availability.

**Table 15. Phone Matches at the Housing Link Level by Form Type**

<b>Form Type</b>	<b>Phone Numbers Matched</b>		<b>Phone Numbers Different</b>		<b>One Number Blank</b>		<b>Both Numbers Blank</b>	
	<b>Number</b>	<b>Percent</b>	<b>Number</b>	<b>Percent</b>	<b>Number</b>	<b>Percent</b>	<b>Number</b>	<b>Percent</b>
MB-MB	180,387	19.0	1,019,944	49.2	193,953	26.5	325,650	81.2
MB-Enum	694,435	73.1	931,322	44.9	458,263	62.5	53,133	13.3
Enum-Enum	74,727	7.9	123,095	5.9	80,594	11.0	22,101	5.5
<b>Total</b>	<b>949,549</b>	<b>100.0</b>	<b>2,074,361</b>	<b>100.0</b>	<b>732,810</b>	<b>100.0</b>	<b>400,884</b>	<b>100.0</b>

Source: Unduplication Analysis Files

Table 15 shows that 73.1 percent of HU to HU links where the phone number matched were enumerated on one side by a Mailout/Mailback response and on the other side by an Enumerator return.

It might be expected that when the whole household (including a single person household) is matching across forms that there is commonality across the returns, and thus it would be more likely that the phone numbers would match. Table 16 shows the results for this expectation. Unlike the previous two tables where non phone matches were shown by ‘Different,’ ‘One Blank,’ and ‘Both Blank,’ they have all been aggregated into ‘Different or Blank’ for Table 16.

**Table 16. Phone Matches at the Housing Link Level by Whole/Partial Status**

<b>Phone Match Status</b>	<b>Whole HH</b>		<b>Single Person HH</b>		<b>Partial HH</b>	
	<b>Number</b>	<b>Percent</b>	<b>Number</b>	<b>Percent</b>	<b>Number</b>	<b>Percent</b>
Phone Numbers Matched	540,954	52.8	216,185	52.4	191,138	7.9
Phone Numbers Different or Blank	484,185	47.2	196,679	47.6	2,226,460	92.1
<b>Total</b>	<b>1,025,139</b>	<b>100.0</b>	<b>412,864</b>	<b>100.0</b>	<b>2,417,598</b>	<b>100.0</b>

Source: Unduplication Analysis Files

Table 16 shows that when whole households or single person households were linked, there was a greater chance of a phone number match (52.8 percent and 52.4 percent, respectively) than for partial households. When the link was a partial household, the phone numbers matched only 7.9 percent of the time between the two responses.

### 5.1.7 GQ Type

Each GQ included on the MAF was assigned a specific GQ Type Code (such as local jails, federal prisons, and federal detention centers). For this evaluation, the type codes were combined into high-level descriptions of GQs (such as correctional facilities) based on the type of service that a GQ provides. Presented in Table 17 are the high level descriptions and the frequency that person links were found between these GQs and a housing unit. For a more detailed description of what each GQ category encompasses and to see which GQ types could claim a UHE<sup>26</sup>, please refer to Appendix A.

<sup>26</sup> Note: Duplication could have been resolved by UHE processing, but the links are retained in these tables to have a complete picture of the type of duplication that occurs.

**Table 17: Type of GQ**

<b>GQ Type</b>	<b>Number</b>	<b>Percent</b>
College/University Student Housing	441,554	51.7
Nursing Facilities/Skilled-Nursing Facilities	149,306	17.5
Correctional Facilities for Adults	104,351	12.2
Soup Kitchens, Transitional Shelters, Mobile Food Vans	39,807	4.7
Other Non-institutional Facilities <sup>27</sup>	36,744	4.3
Group Homes and Residential Treatment Centers Intended for Adults	20,963	2.5
Military Quarters	17,689	2.1
Juvenile Facilities	17,675	2.1
Unknown GQ Type	13,890	1.6
Other Institutional Facilities <sup>28</sup>	11,977	1.4
<b>Total</b>	<b>853,956</b>	<b>100.0</b>

Source: Unduplication Analysis Files

Over half of all links found between a housing unit and a GQ involved college or university housing (51.7 percent). The next most common type of GQ where duplicates were enumerated was nursing facilities (17.5 percent), followed by correctional facilities for adults (12.2 percent).<sup>29</sup>

Table 3 previously showed that HU to GQ links were most commonly found within the same state. Table 18 now shows how often HU to GQ links occurred by geographic proximity for each GQ type.

<sup>27</sup> Includes living quarters for victims of natural disasters, U.S. flag maritime/merchant vessels, religious group quarters, workers' group living quarters, and Job Corps Centers.

<sup>28</sup> Includes mental hospitals, hospitals with patients who have no usual home elsewhere, in-patient hospice facilities, military treatment facilities, and residential schools for people with disabilities

<sup>29</sup> Results from the enumeration of Group Quarters in 2010 were not available at the time of this writing, but results from 2000 showed that colleges and universities constituted 26.4 percent, correctional facilities constituted 25.5 percent, and nursing homes accounted for 22.1 percent of the entire GQ population.

**Table 18. Number and Percent of GQ Types by Geography**

<b>GQ Type</b>	<b>Overall</b>		<b>Within Block</b>		<b>Surrounding Block</b>		<b>Within County</b>		<b>Within State</b>		<b>Across State</b>	
	<b>Number</b>	<b>Percent</b>	<b>Number</b>	<b>Percent</b>	<b>Number</b>	<b>Percent</b>	<b>Number</b>	<b>Percent</b>	<b>Number</b>	<b>Percent</b>	<b>Number</b>	<b>Percent</b>
College/University Student Housing	<b>441,554</b>	<b>100.0</b>	13,389	3.0	2,535	0.6	49,482	11.2	267,301	60.5	108,847	24.7
Nursing or Skilled Nursing Facilities	<b>149,306</b>	<b>100.0</b>	39,223	26.3	4,410	3.0	81,177	54.4	21,070	14.1	3,426	2.3
Correctional Facilities for Adults	<b>104,351</b>	<b>100.0</b>	188	0.2	500	0.5	56,963	54.6	38,726	37.1	7,974	7.6
Soup Kitchens, Transitional Shelters, Mobile Food Vans	<b>39,807</b>	<b>100.0</b>	5,130	12.9	2,125	5.3	27,832	69.9	3,436	8.6	1,284	3.2
Other Non-institutional Facilities	<b>36,744</b>	<b>100.0</b>	21,012	57.2	1,314	3.6	3,163	8.6	6,439	17.5	4,816	13.1
Group Homes and Residential Treatment Centers Intended for Adults	<b>20,963</b>	<b>100.0</b>	5,416	25.8	851	4.1	7,537	36.0	5,411	25.8	1,748	8.3
Military Quarters	<b>17,689</b>	<b>100.0</b>	183	1.0	150	0.8	3,886	22.0	2,516	14.2	10,954	61.9
Juvenile Facilities	<b>17,675</b>	<b>100.0</b>	171	1.0	108	0.6	6,488	36.7	9,608	54.4	1,300	7.4
Unknown GQ Type	<b>13,890</b>	<b>100.0</b>	8,642	62.2	508	3.7	1,832	13.2	2,207	15.9	701	5.0
Other Institutional Facilities	<b>11,977</b>	<b>100.0</b>	392	3.3	140	1.2	5,433	45.4	5,236	43.7	776	6.5
<b>Total</b>	<b>853,956</b>	<b>100.0</b>	93,746	11.0	12,641	1.5	243,793	28.5	361,950	42.4	141,826	16.6

Source: Unduplication Analysis Files

GQ Types that linked to a housing unit within the same county more often than any other geography included:

- Soup kitchens, transitional shelters, and mobile food vans, 69.9 percent within county



- Correctional facilities for adults, 54.6 percent within county
- Nursing facilities/skilled nursing facilities, 54.4 percent within county
- Other institutional facilities (such as mental hospitals and hospitals with patients with no UHE), 45.4 percent within county

GQ Types that linked to a housing unit within the same state more often than any other geography included:

- College/University Student Housing, 60.5 percent within state
- Juvenile Facilities, 54.4 percent within state
- Other institutional Facilities (such as mental hospitals and hospitals with patients with no UHE), 43.7 percent within state

Military quarters were the only GQ type to have over half of the duplication (61.9 percent) occurring across state lines. Non-institutional and unknown GQ Types frequently involved duplication within the same block (57.2 percent and 62.2 percent, respectively), which could indicate that those GQs were included on the MAF both with an HU record and with a GQ record. Additionally, misclassification could also have happened for Nursing Facilities and Group Home GQs, causing within block duplication (which occurred at 26.3 percent and 25.8 percent, respectively, within those categories).

### 5.1.8 United States and Puerto Rico

Links could have been identified between any of the fifty states, the District of Columbia, and Puerto Rico. Table 19 shows the frequency at which two responses were linked with both addresses stateside and both addresses in Puerto Rico. Matching between stateside and Puerto Rico was considered out of scope (Lynch, 2009a).

**Table 19: Frequency of Matching Between Stateside Addresses and Puerto Rico Addresses**

Match Type	Response Level		Person Level	
	Number	Percent	Number	Percent
Both Addresses Stateside	4,634,739	98.4	7,282,347	97.7
Both Addresses in Puerto Rico	76,821	1.6	171,824	2.3
<b>Total</b>	<b>4,711,560</b>	<b>100.0</b>	<b>7,454,171</b>	<b>100.0</b>

Source: Unduplication Analysis Files

Table 19 shows that the majority of response links in this evaluation (98.4 percent) were between two stateside addresses.

Table 20 shows the distribution of person links for both stateside cases (US) and Puerto Rico (PR).

**Table 20. US and Puerto Rico Links at the Person Link Level, by Geography**

Geography	US or PR Link					
	Overall		US		PR	
	Number	Percent	Number	Percent	Number	Percent
Within Block	2,495,776	33.5	2,388,902	32.8	106,874	62.2
Surrounding Block	786,273	10.5	746,777	10.3	39,496	23.0
Within County	2,085,442	28.0	2,073,287	28.5	12,155	7.1
Within State	1,304,804	17.5	1,291,505	17.7	13,299	7.7
Across State	781,876	10.5	781,876	10.7	0	0.0
<b>Total</b>	<b>7,454,171</b>	<b>100.0</b>	<b>7,282,347</b>	<b>100.0</b>	<b>171,824</b>	<b>100.0</b>

Source: Unduplication Analysis Files

When looking at Puerto Rico, we see that 62.2 percent of person links are found within the block and 23.0 percent of person links are found in surrounding blocks. These percentages are much higher than the stateside counterparts, indicating that address issues might be the primary cause of duplication in Puerto Rico.

Due to the limitation of not matching between the U.S. and Puerto Rico, any comparisons should be made with caution.

### 5.1.9 Overcount Question

Each person rostered in the census on a housing unit questionnaire could indicate through the overcount question whether they lived or stayed at another location. The overcount question that was printed on the mail form was shown earlier in Figure 2.

In the subsequent tables, we refer to the possible overcount categories that could have been marked (such as custody or jail). We also refer to ‘Other below the line’ cases’, which are cases that marked either the ‘for another reason’ box, or simply marked ‘yes’ without specifying a category. ‘Other above the line cases’ are responses identified as having multiple overcount marks within the household or within a person record. They could also have been responses from TQA that utilized the combined nursing home/jail category. These are discussed in more detail in Section 5.2.

Since the links in this evaluation represent duplicated persons who seem to have more than one place where they lived or stayed, it would be natural for a high percentage of these links to have positively marked the overcount question on at least one of the HU to HU responses, or on the housing unit side of a HU-GQ link. The tables below indicate how often at least one or multiple persons on a housing unit response indicated an overcount reason; the positive overcount mark might not necessarily have been associated with the duplicated person. Table 21 shows how often HU to HU person links marked certain HU-related categories in the overcount question on either housing unit’s return.

**Table 21: Overcount Indication Provided By HU to HU Links**

<b>Overcount Question Response</b>	<b>Number</b>	<b>Percent</b>
No overcount mark	3,876,009	58.7
Both seasonal	381,154	5.8
One side seasonal	429,836	6.5
Both custody	317,489	4.8
One side Custody	295,551	4.5
Other “Below the line” marks <sup>30</sup> /Blank	447,247	6.8
Other markings	852,929	12.9
<b>Total</b>	<b>6,600,215</b>	<b>100.0</b>

Source: Unduplication Analysis Files

Table 21 shows that 58.7 percent of HU-HU person links did not have anyone in either housing unit who positively marked the overcount box. Over twelve percent of HU to HU person links had the seasonal box marked on one or both of the two linked returns. Over nine percent had the custody box marked on one or both of the two linked returns. Those cases with other markings of the overcount question could have marked multiple boxes, marked the ‘for another reason’ box, marked ‘yes’ without indicating a category, or marked the supposedly GQ-related categories.

Table 22 shows how often HU-GQ person links marked certain GQ-related categories in the overcount question on the housing unit response.

**Table 22: Overcount Indication Provided By HU to GQ Links**

<b>Overcount Question Response</b>	<b>Number</b>	<b>Percent</b>
No overcount mark	352,324	41.3
College	331,543	38.8
Nursing Home	30,078	3.5
Jail	20,344	2.4
Military	10,653	1.2
Other “Above the line” marks <sup>31</sup>	56,655	6.6
Other “Below the line” marks	35,245	4.1
Other markings	17,114	2.0
<b>Total</b>	<b>853,956</b>	<b>100.0</b>

Source: Unduplication Analysis Files

Table 22 shows that over forty percent of HU-GQ links did not give an indication in the overcount question of having another place where they live or stay. Of the links that did positively mark the overcount question, 38.8 percent indicated they stayed elsewhere for college. Those cases with other markings of the overcount question either marked ‘yes’ without indicating a category, or marked the supposedly housing unit related categories of seasonal or custody (as opposed to one of the GQ-related categories in the table).

<sup>30</sup> Includes the Overcount categories “Yes only” and “Another Reason”.

<sup>31</sup> Includes the Overcount categories “Person Multiple” and “Household Multiple”.

Table 23 shows the occurrence of an overcount mark on responses for all HU-HU person links, distributed by geography.

**Table 23: Presence of Overcount Mark for HU to HU Person Links, By Geography**

<b>Geography</b>	<b>Total HU to HU Person Links</b>	<b>No overcount mark</b>		<b>With overcount mark</b>	
	<b>Number</b>	<b>Number</b>	<b>Percent</b>	<b>Number</b>	<b>Percent</b>
Within Block	2,402,030	2,019,646	84.1	382,384	15.9
Surrounding Block	773,632	625,955	80.9	147,677	19.1
Within County	1,841,649	815,551	44.3	1,026,098	55.7
Within State	942,854	260,598	27.6	682,256	72.4
Across State	640,050	154,259	24.1	485,791	75.9
<b>Total</b>	<b>6,600,215</b>	<b>3,876,009</b>	<b>58.7</b>	<b>2,724,206</b>	<b>41.3</b>

Source: Unduplication Analysis Files

Links within the same block or within surrounding blocks did not often include a positive overcount mark; 84.1 percent of within block HU-HU person links had no overcount mark, and 80.9 percent of surrounding block HU-HU person links had no overcount mark. Since these duplications were likely caused by address issues, it is expected that they would not mark the overcount box, since the overcount question pertains to complex living situations and person-level duplication issues. The overcount question was marked more often for duplication across longer distances; only 24.1 percent of links across state lines and 27.6 percent of links within the same state did not mark the overcount question. This suggests that future resolution of person-level long-distance duplication cases could come through additional probing, after the overcount question is flagged on an initial census return.

Table 24 shows the occurrence of an overcount mark on responses for all HU-GQ person links, distributed by geography.

**Table 24: Presence of Overcount Mark for HU to GQ Person Links, By Geography**

<b>Geography</b>	<b>Total HU to GQ Person Links</b>	<b>No overcount mark</b>		<b>With overcount mark</b>	
	<b>Number</b>	<b>Number</b>	<b>Percent</b>	<b>Number</b>	<b>Percent</b>
Within Block	93,746	80,704	86.1	13,042	13.9
Surrounding Block	12,641	9,300	73.6	3,341	26.4
Within County	243,793	147,542	60.5	96,251	39.5
Within State	361,950	86,537	23.9	275,413	76.1
Across State	141,826	28,241	19.9	113,585	80.1
<b>Total</b>	<b>853,956</b>	<b>352,324</b>	<b>41.3</b>	<b>501,632</b>	<b>58.7</b>

Source: Unduplication Analysis Files

The results in Table 24 are similar to those from Table 23. HU-GQ links within the same block or within surrounding blocks did not often include a positive overcount mark; 86.1 percent of within block HU-GQ person links had no overcount mark, and 73.6 percent of surrounding block HU-HU person links had no overcount mark. The overcount question was marked more often for duplication across longer distances; only 23.9 percent of links within state lines and 19.9

percent of links across state lines did not mark the overcount question. However, 60.5 percent of within county HU-GQ links did not mark the overcount question, a higher rate than observed in HU-HU links (44.3 percent). This could be a result of the saliency of these living situations; within county HU-GQ links might have originated from short GQ stays, such as a brief period in a local jail or a few weeks in a nursing home for rehabilitation, which were not permanent enough living situations to prompt an overcount indication.

### **5.1.10 Demographic Characteristics of Linked Persons**

There were 7,454,171 person links included on 4,711,560 response links in the 2010 Census. This section will present the demographic characteristics for these person links. Table 25 to Table 35 report demographic characteristics of the potential duplicate persons: age, Hispanic origin, race, and sex. Age was calculated based on the date of birth provided; if no date of birth was provided then the write-in age was used. Age was calculated only if the date of birth fell within valid date ranges. Similarly, the calculated age or write-in age was used only if it fell within valid age ranges; otherwise it was considered missing.<sup>32</sup>

Since this evaluation looks at person links and not a single person, the information provided on each side of the link was compared to the other side. It was possible for the sides of a link to provide different information on the duplicated person or for one side to provide information and the other to not provide any information, even though we considered them to be a match. For instance, one side might have listed a child as being 9 years old and the other side listed the child as being 10 years old. The following tables show how often the two sides agreed on the demographic characteristics. If one side was blank, then the person was categorized using the information provided by the other side (assuming it was nonblank).

The demographic data used in this assessment are unedited so there is a row in the following tables for people with missing values for the specific characteristic. Direct comparisons with published 2010 Census results are not possible since the data in published Census reports have undergone editing and imputation, and therefore will have no missing values.

---

<sup>32</sup> This is a different calculated age than was used for the matching process.

## Age

Table 25 shows how often the age category (shown in Table 28) either agreed between the two sides of a link, disagreed between the two sides, was blank on one side, or was blank on both sides.

**Table 25. Age Category Match Status By Geography**

Age Category Match Status	Overall		Within Block		Surrounding Block		Within County		Within State		Across State	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent
Agree	7,009,995	94.0	2,175,777	87.2	746,373	94.9	2,031,844	97.4	1,283,719	98.4	772,282	98.8
Disagree	171,474	2.3	92,660	3.7	20,180	2.6	37,309	1.8	15,584	1.2	5,741	0.7
One Side Blank	260,728	3.5	216,525	8.7	19,141	2.4	15,933	0.8	5,378	0.4	3,751	0.5
Both Blank	11,974	0.2	10,814	0.4	579	0.1	356	0.0	123	0.0	102	0.0
<b>Total</b>	<b>7,454,171</b>	<b>100.0</b>	<b>2,495,776</b>	<b>100.0</b>	<b>786,273</b>	<b>100.0</b>	<b>2,085,442</b>	<b>100.0</b>	<b>1,304,804</b>	<b>100.0</b>	<b>781,876</b>	<b>100.0</b>

Source: Unduplication Analysis Files

Besides a person's name, the age variable was the most heavily weighted demographic item used in matching. Thus it is not surprising for Table 25 to show that, overall, 94.0 percent of person links provided a calculated age that agreed between the two sides (agreement means that the provided ages were in the same category as detailed in the next table). Thus, an age of 9 on one side would be considered to be in agreement with a reported age of 8 on the other side, but an age of 9 on one side would not agree with a reported age of 10 on the other side. It should be noted that the actual matching process assigned a match score for the age match using a prorated score that allowed for some differences in age depending on the magnitude of the difference and the calculated age. In Table 25, within-block matches were the only time the rate of agreement fell below 90 percent, but that is impacted by 8.7 percent of within block links that only had age reported on one side of the link. This is not surprising since the matching process was more lenient for within block links.

Table 26 shows the ages of the persons identified as potential duplicates overall and within each geography. The row of 'Both blank' from Table 25 appears as 'Missing' in Table 26. The row of 'Disagree' from Table 25 appears as 'Inconsistent' in Table 26. All links classified as either 'Agree' or 'One Side Blank' in Table 25 are distributed across the age categories in Table 26.

**Table 26. Age of Duplicates by Geography**

<b>Age in Years</b>	<b>Overall</b>		<b>Within Block</b>		<b>Surrounding Block</b>		<b>Within County</b>		<b>Within State</b>		<b>Across State</b>	
	<b>Number</b>	<b>Percent</b>	<b>Number</b>	<b>Percent</b>	<b>Number</b>	<b>Percent</b>	<b>Number</b>	<b>Percent</b>	<b>Number</b>	<b>Percent</b>	<b>Number</b>	<b>Percent</b>
Under 5 years	447,959	6.0	132,630	5.3	49,909	6.3	180,563	8.7	59,546	4.6	25,311	3.2
5 to 9 years	526,291	7.1	125,574	5.0	53,720	6.8	252,213	12.1	73,062	5.6	21,722	2.8
10 to 14 years	580,662	7.8	131,147	5.3	57,206	7.3	284,216	13.6	82,664	6.3	25,429	3.3
15 to 19 years	853,916	11.5	150,845	6.0	56,252	7.2	288,243	13.8	262,026	20.1	96,550	12.3
20 to 24 years	878,278	11.8	179,224	7.2	50,756	6.5	231,782	11.1	298,086	22.8	118,430	15.1
25 to 29 years	465,653	6.2	166,900	6.7	50,655	6.4	144,933	6.9	68,646	5.3	34,519	4.4
30 to 34 years	353,585	4.7	146,262	5.9	46,459	5.9	96,720	4.6	41,479	3.2	22,665	2.9
35 to 39 years	325,561	4.4	142,433	5.7	45,296	5.8	79,684	3.8	35,955	2.8	22,193	2.8
40 to 44 years	335,884	4.5	151,524	6.1	45,912	5.8	73,795	3.5	38,938	3.0	25,715	3.3
45 to 49 years	383,374	5.1	173,061	6.9	51,762	6.6	75,068	3.6	48,142	3.7	35,341	4.5
50 to 54 years	399,302	5.4	178,110	7.1	51,881	6.6	68,424	3.3	55,939	4.3	44,948	5.7
55 to 59 years	378,233	5.1	161,948	6.5	47,749	6.1	56,532	2.7	58,438	4.5	53,566	6.9
60 to 64 years	354,145	4.8	143,832	5.8	43,487	5.5	45,836	2.2	55,337	4.2	65,653	8.4
65 to 69 years	280,943	3.8	108,668	4.4	34,263	4.4	34,839	1.7	40,146	3.1	63,027	8.1
70 to 74 years	212,999	2.9	82,306	3.3	26,456	3.4	29,094	1.4	25,755	2.0	49,388	6.3
75 to 79 years	170,811	2.3	68,125	2.7	20,865	2.7	28,868	1.4	17,880	1.4	35,073	4.5
80 years and over	323,127	4.3	149,713	6.0	32,886	4.2	76,967	3.7	27,058	2.1	36,503	4.7
Inconsistent	171,474	2.3	92,660	3.7	20,180	2.6	37,309	1.8	15,584	1.2	5,741	0.7
Missing	11,974	0.2	10,814	0.4	579	0.1	356	0.0	123	0.0	102	0.0
<b>Total</b>	<b>7,454,171</b>	<b>100.0</b>	<b>2,495,776</b>	<b>100.0</b>	<b>786,273</b>	<b>100.0</b>	<b>2,085,442</b>	<b>100.0</b>	<b>1,304,804</b>	<b>100.0</b>	<b>781,876</b>	<b>100.0</b>

Source: Unduplication Analysis Files

Of the persons that were identified as potential duplicates, 11.8 percent were 20 to 24 years of age and 11.5 percent were persons 15 to 19 years of age.

Table 27 presents one column for the 7,454,171 million duplicates found in the census (also presented in Table 26) and a second column for the 308,745,538 person records enumerated in the final 2010 Census records. The data on the duplicates are taken directly from the questionnaire without imputation or editing while the final census results have been cleaned, so there are no inconsistent or missing data for that column.

**Table 27. Age of Duplicates Compared to Age of All Enumerated Persons**

Age in Years	Duplicates Found in the Census		All Persons Enumerated in the Census	
	Number	Percent	Number	Percent
Under 5 years	447,959	6.0	20,201,362	6.5
5 to 9 years	526,291	7.1	20,348,657	6.6
10 to 14 years	580,662	7.8	20,677,194	6.7
15 to 19 years	853,916	11.5	22,040,343	7.1
20 to 24 years	878,278	11.8	21,585,999	7.0
25 to 29 years	465,653	6.2	21,101,849	6.8
30 to 34 years	353,585	4.7	19,962,099	6.5
35 to 39 years	325,561	4.4	20,179,642	6.5
40 to 44 years	335,884	4.5	20,890,964	6.8
45 to 49 years	383,374	5.1	22,708,591	7.4
50 to 54 years	399,302	5.4	22,298,125	7.2
55 to 59 years	378,233	5.1	19,664,805	6.4
60 to 64 years	354,145	4.8	16,817,924	5.4
65 to 69 years	280,943	3.8	12,435,263	4.0
70 to 74 years	212,999	2.9	9,278,166	3.0
75 to 79 years	170,811	2.3	7,317,795	2.4
80 years and over	323,127	4.3	11,236,760	3.6
Inconsistent	171,474	2.3	NA	NA
Missing	11,974	0.2	NA	NA
<b>Total</b>	<b>7,454,171</b>	<b>100.0</b>	<b>308,745,538</b>	<b>100.0</b>

Source: Unduplication Analysis Files



Duplicates show up with particular frequency in 15 to 24 year olds. That age cohort makes up only about seven percent of the total population but over eleven percent of all duplicates. Children aged five to fourteen also occur with greater frequency in the universe of duplicates than the overall population, as do persons 80 years old and over.

### *Hispanic Origin*

Table 20 shows how often Hispanic origin agreed between the two sides of a link, disagreed between the two sides, was blank on one side, or was blank on both sides. Hispanic origin is not included in the matching process because it is a demographic item that can vary depending if the data are being reported by a household member or a proxy, thus it cannot be used to reliably increase or decrease the chances of a person match.

**Table 28. Hispanic Origin Category Match Status By Geography**

Hispanic Origin	Overall		Within Block		Surrounding Block		Within County		Within State		Across State	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent
Agree	6,150,836	82.5	2,067,348	82.8	691,355	87.9	1,672,143	80.2	1,059,911	81.2	660,079	84.4
Disagree	417,067	5.6	161,429	6.5	38,050	4.8	137,394	6.6	52,552	4.0	27,642	3.5
One Side Blank	828,277	11.1	245,970	9.9	53,428	6.8	259,391	12.4	181,086	13.9	88,402	11.3
Both Blank	57,991	0.8	21,029	0.8	3,440	0.4	16,514	0.8	11,255	0.9	5,753	0.7
<b>Total</b>	<b>7,454,171</b>	<b>100.0</b>	<b>2,495,776</b>	<b>100.0</b>	<b>786,273</b>	<b>100.0</b>	<b>2,085,442</b>	<b>100.0</b>	<b>1,304,804</b>	<b>100.0</b>	<b>781,876</b>	<b>100.0</b>

Source: Unduplication Analysis Files

Table 28 helps confirm the decision that Hispanic origin should not be used as a matching variable, since the reported characteristic only matches 82.5 percent of the time.

Table 29 shows the Hispanic origin of the persons identified as potential duplicate persons overall and within each geography.

**Table 29. Hispanic Origin of Duplicates by Geography**

<b>Hispanic Origin</b>	<b>Overall</b>		<b>Within Block</b>		<b>Surrounding Block</b>		<b>Within County</b>		<b>Within State</b>		<b>Across State</b>	
	<b>Number</b>	<b>Percent</b>	<b>Number</b>	<b>Percent</b>	<b>Number</b>	<b>Percent</b>	<b>Number</b>	<b>Percent</b>	<b>Number</b>	<b>Percent</b>	<b>Number</b>	<b>Percent</b>
Not Hispanic or Latino checkbox only	6,019,444	80.8	1,875,189	75.1	626,989	79.7	1,664,357	79.8	1,141,318	87.5	711,591	91.0
Mexican checkbox only	472,696	6.3	184,808	7.4	48,146	6.1	169,298	8.1	52,164	4.0	18,280	2.3
Puerto Rican checkbox only	260,395	3.5	146,779	5.9	46,703	5.9	38,955	1.9	22,484	1.7	5,474	0.7
Cuban checkbox only	35,990	0.5	14,902	0.6	4,287	0.5	11,536	0.6	3,486	0.3	1,779	0.2
Another Hispanic checkbox only	5,326	0.1	1,891	0.1	428	0.1	1,885	0.1	821	0.1	301	0.0
Multiple checkboxes	3,333	0.0	1,036	0.0	235	0.0	1,237	0.1	554	0.0	271	0.0
Both Checkbox and Write-in	171,921	2.3	83,941	3.4	17,328	2.2	41,651	2.0	18,931	1.5	10,070	1.3
Write-in Only	10,008	0.1	4,772	0.2	667	0.1	2,615	0.1	1,239	0.1	715	0.1
Inconsistent	417,067	5.6	161,429	6.5	38,050	4.8	137,394	6.6	52,552	4.0	27,642	3.5
Missing	57,991	0.8	21,029	0.8	3,440	0.4	16,514	0.8	11,255	0.9	5,753	0.7
<b>Total</b>	<b>7,454,171</b>	<b>100.0</b>	<b>2,495,776</b>	<b>100.0</b>	<b>786,273</b>	<b>100.0</b>	<b>2,085,442</b>	<b>100.0</b>	<b>1,304,804</b>	<b>100.0</b>	<b>781,876</b>	<b>100.0</b>

Source: Unduplication Analysis Files

Of the persons identified as duplicates, 80.8 percent were not of Hispanic or Latino origin. About 75.1 percent of within block person links were not of Hispanic or Latino origin and this percentage increased across geographies, until it reached a high of 91.0 percent at the across state geography.

Table 30 presents one column with the distribution of Hispanic Origin for the 7,454,171 million duplicates found in the census (also shown in Table 29) and adds a second column for the 308,745,538 person records enumerated in the final census records. The data on the duplicates are taken directly from the questionnaire without imputation or editing while the final census results have been cleaned, so there are no missing or inconsistent data for that column as well as no explicit identification of multiple origins. Data analysis on duplicates was unable to consider write-in fields when assigning a Hispanic Origin. Thus, duplicates identified as Mexican (or Cuban, or Puerto Rican) in the following table marked the checkbox for Mexican (or Cuban, or Puerto Rican) and did not write anything in to

the provided fields, while those included as Mexicans (or Cubans, or Puerto Ricans) from the entire universe could have utilized the write-in and thus been coded or edited into another category.

**Table 30. Hispanic Origin of Duplicates Compared to All Enumerated Persons**

<b>Hispanic Origin</b>	<b>Duplicates Found in the Census</b>		<b>All Persons Enumerated in the Census</b>	
	<b>Number</b>	<b>Percent</b>	<b>Number</b>	<b>Percent</b>
Not Hispanic or Latino	6,019,444	80.8	258,267,944	83.7
Mexican	472,696	6.3	31,798,258	10.3
Puerto Rican	260,395	3.5	4,623,716	1.5
Cuban	35,990	0.5	1,785,547	0.6
Another Hispanic Origin	5,326	0.1	12,270,073	4.0
Multiple checkboxes	3,333	0.0	NA	NA
Both Checkbox and Write-in	171,921	2.3	NA	NA
Write-in Only	10,008	0.1	NA	NA
Inconsistent	417,067	5.6	NA	NA
Missing	57,991	0.8	NA	NA
<b>Total</b>	<b>7,454,171</b>	<b>100.0</b>	<b>308,745,538</b>	<b>100.0</b>

Source: Unduplication Analysis Files

Using different classification systems as described before Table 30, 83.7 percent of the country was identified as non-Hispanic in the 2010 Census, compared to 80.8 percent of duplicates classified as non-Hispanic. However, 5.6 percent of duplicates were classified as having inconsistent markings to this question across the two linked responses.

### ***Race***

Table 31 shows how often the race category (defined in Table 32) agreed or disagreed between each side of a link, or how often one side was blank, or how often both sides were blank. Race is not included in the matching process because it is a demographic item that can vary depending on the respondent. Thus it cannot be used to reliably increase or decrease the chances of a person match.

**Table 31. Race Category Match Status By Geography**

<b>Race</b>	<b>Overall</b>		<b>Within Block</b>		<b>Surrounding Block</b>		<b>Within County</b>		<b>Within State</b>		<b>Across State</b>	
	<b>Number</b>	<b>Percent</b>	<b>Number</b>	<b>Percent</b>	<b>Number</b>	<b>Percent</b>	<b>Number</b>	<b>Percent</b>	<b>Number</b>	<b>Percent</b>	<b>Number</b>	<b>Percent</b>
Agree	5,932,871	79.6	1,946,737	78.0	665,010	84.6	1,621,509	77.8	1,047,535	80.3	652,080	83.4
Disagree	835,225	11.2	328,573	13.2	79,798	10.1	245,669	11.8	114,996	8.8	66,189	8.5
One Side												
Blank	639,279	8.6	202,190	8.1	38,561	4.9	203,260	9.7	135,105	10.4	60,163	7.7
Both Blank	46,796	0.6	18,276	0.7	2,904	0.4	15,004	0.7	7,168	0.5	3,444	0.4
<b>Total</b>	<b>7,454,171</b>	<b>100.0</b>	<b>2,495,776</b>	<b>100.0</b>	<b>786,273</b>	<b>100.0</b>	<b>2,085,442</b>	<b>100.0</b>	<b>1,304,804</b>	<b>100.0</b>	<b>781,876</b>	<b>100.0</b>

Source: Unduplication Analysis Files

Table 31 helps confirm the decision that race should not be used as a matching variable, since it only matched 79.6 percent of the time and disagreed 11.2 percent of the time.

Table 32 shows the race of the persons identified as duplicates overall and within each geography.

Table 32. Race of Duplicates by Geography

Race	Overall		Within Block		Surrounding Block		Within County		Within State		Across State	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent
White checkbox alone	4,885,772	65.5	1,542,868	61.8	535,844	68.1	1,293,121	62.0	915,738	70.2	598,201	76.5
Black or African American checkbox alone	1,012,163	13.6	340,361	13.6	104,580	13.3	346,968	16.6	157,715	12.1	62,539	8.0
American Indian and Alaska Native checkbox alone	4,038	0.1	1,613	0.1	379	0.0	1,274	0.1	546	0.0	226	0.0
Asian Indian checkbox alone	42,901	0.6	13,639	0.5	5,799	0.7	7,334	0.4	9,571	0.7	6,558	0.8
Chinese checkbox alone	76,564	1.0	29,044	1.2	4,137	0.5	15,017	0.7	19,419	1.5	8,947	1.1
Filipino checkbox alone	43,446	0.6	15,956	0.6	3,754	0.5	11,252	0.5	8,683	0.7	3,801	0.5
Japanese checkbox alone	11,652	0.2	4,872	0.2	1,031	0.1	3,083	0.1	1,602	0.1	1,064	0.1
Korean checkbox alone	28,073	0.4	9,009	0.4	2,697	0.3	6,522	0.3	6,574	0.5	3,271	0.4
Vietnamese checkbox alone	23,976	0.3	6,724	0.3	1,881	0.2	7,526	0.4	5,854	0.4	1,991	0.3
Native Hawaiian checkbox alone	1,594	0.0	662	0.0	203	0.0	473	0.0	156	0.0	100	0.0
Guamanian or Chamorro checkbox alone	806	0.0	291	0.0	96	0.0	249	0.0	93	0.0	77	0.0
Samoan checkbox alone	973	0.0	336	0.0	95	0.0	348	0.0	121	0.0	73	0.0
Other Asian checkbox alone	289	0.0	106	0.0	14	0.0	74	0.0	64	0.0	31	0.0
Other Pacific Islander checkbox alone	74	0.0	28	0.0	3	0.0	24	0.0	16	0.0	3	0.0
Some Other Race checkbox alone	3,504	0.0	1,721	0.1	299	0.0	1,075	0.1	308	0.0	101	0.0
Multiple checkboxes	55,303	0.7	14,731	0.6	5,071	0.6	21,230	1.0	10,069	0.8	4,202	0.5
Both Checkbox and Write-in	355,072	4.8	156,836	6.3	36,077	4.6	100,263	4.8	42,614	3.3	19,282	2.5
Write-in Only	25,950	0.3	10,130	0.4	1,611	0.2	8,936	0.4	3,497	0.3	1,776	0.2
Inconsistent	835,225	11.2	328,573	13.2	79,798	10.1	245,669	11.8	114,996	8.8	66,189	8.5
Missing	46,796	0.6	18,276	0.7	2,904	0.4	15,004	0.7	7,168	0.5	3,444	0.4
<b>Total</b>	<b>7,454,171</b>	<b>100.0</b>	<b>2,495,776</b>	<b>100.0</b>	<b>786,273</b>	<b>100.0</b>	<b>2,085,442</b>	<b>100.0</b>	<b>1,304,804</b>	<b>100.0</b>	<b>781,876</b>	<b>100.0</b>

Source: Unduplication Analysis Files

Of the persons identified as duplicates, 65.5 percent marked the White checkbox alone, 13.6 percent marked the Black or African American checkbox alone, and 11.2 percent provided contradicting race categories. All other race categories were fairly small.

Table 33 presents one column with the distribution of race for the 7,454,171 million duplicates found in the census (also shown in Table 32) and adds a second column for the 308,745,538 person records enumerated in the final 2010 Census records. The data on the duplicates are taken directly from the questionnaire without imputation or editing while the final census results have been cleaned, so there are no missing or inconsistent data for that column. As with Hispanic Origin, data analysis on the duplicates was unable to consider write-in fields when assigning a race.

**Table 33. Race of Duplicates Compared to All Enumerated Persons**

<b>Race</b>	<b>Duplicates Found in the Census</b>		<b>All Persons Enumerated in the Census</b>	
	<b>Number</b>	<b>Percent</b>	<b>Number</b>	<b>Percent</b>
White	4,885,772	65.5	223,553,265	72.4
Black or African American	1,012,163	13.6	38,929,319	12.6
American Indian and Alaska Native	4,038	0.1	2,932,248	0.9
Asian Indian	42,901	0.6	2,843,391	0.9
Chinese	76,564	1.0	3,347,229	1.1
Filipino	43,446	0.6	2,555,923	0.8
Japanese	11,652	0.2	763,325	0.2
Korean	28,073	0.4	1,423,784	0.5
Vietnamese	23,976	0.3	1,548,449	0.5
Native Hawaiian	1,594	0.0	156,146	0.1
Guamanian or Chamorro	806	0.0	88,310	0.0
Samoan	973	0.0	109,637	0.0
Other Asian	289	0.0	2,192,151	0.7
Other Pacific Islander	74	0.0	185,920	0.1
Some Other Race	3,504	0.0	19,107,368	6.2
Multiple checkboxes	55,303	0.7	9,009,073	2.9
Both Checkbox and Write-in	355,072	4.8	NA	NA
Write-in Only	25,950	0.3	NA	NA
Inconsistent	835,225	11.2	NA	NA
Missing	46,796	0.6	NA	NA
<b>Total</b>	<b>7,454,171</b>	<b>100.0</b>	<b>308,745,538</b>	<b>100.0</b>

Source: Unduplication Analysis Files

Using different classification systems as described before Table 33, 72.4 percent of the country was identified as White in the 2010 Census, compared to 65.5 percent of duplicates who identified as White. However, 11.2 percent of duplicates were classified as having inconsistent markings to this question across the two linked responses.

## Sex

Table 34 shows how often the sex category that each person identified as a potential duplicate agrees or disagrees; or how often one side is blank or both sides are blank. Sex is one of the key variables used in the matching process.

**Table 34. Sex Match Status By Geography**

Sex	Overall		Within Block		Surrounding Block		Within County		Within State		Across State	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent
Agree	7,167,917	96.2	2,380,162	95.4	755,310	96.1	2,002,563	96.0	1,266,832	97.1	763,050	97.6
Disagree	50,269	0.7	26,767	1.1	7,953	1.0	9,932	0.5	3,156	0.2	2,461	0.3
One Side Blank	231,799	3.1	87,110	3.5	22,672	2.9	71,800	3.4	34,182	2.6	16,035	2.1
Both Blank	4,186	0.1	1,737	0.1	338	0.0	1,147	0.1	634	0.0	330	0.0
<b>Total</b>	<b>7,454,171</b>	<b>100.0</b>	<b>2,495,776</b>	<b>100.0</b>	<b>786,273</b>	<b>100.0</b>	<b>2,085,442</b>	<b>100.0</b>	<b>1,304,804</b>	<b>100.0</b>	<b>781,876</b>	<b>100.0</b>

Source: Unduplication Analysis Files

Of all person links, 96.2 percent agreed on describing a person's sex. The information was missing from one side an additional 3.1 percent of the time.

Table 35 shows the sex of the persons identified as potential duplicate persons overall and within each geography.



**Table 35. Sex of Duplicates by Geography**

<b>Sex</b>	<b>Overall</b>		<b>Within Block</b>		<b>Surrounding Block</b>		<b>Within County</b>		<b>Within State</b>		<b>Across State</b>	
	<b>Number</b>	<b>Percent</b>	<b>Number</b>	<b>Percent</b>	<b>Number</b>	<b>Percent</b>	<b>Number</b>	<b>Percent</b>	<b>Number</b>	<b>Percent</b>	<b>Number</b>	<b>Percent</b>
Male	3,725,233	50.0	1,202,814	48.2	378,337	48.1	1,070,100	51.3	669,161	51.3	404,821	51.8
Female	3,674,393	49.3	1,264,411	50.7	399,637	50.8	1,004,245	48.2	631,842	48.4	374,258	47.9
Both	90	0.0	47	0.0	8	0.0	18	0.0	11	0.0	6	0.0
Inconsistent	50,269	0.7	26,767	1.1	7,953	1.0	9,932	0.5	3,156	0.2	2,461	0.3
Missing	4,186	0.1	1,737	0.1	338	0.0	1,147	0.1	634	0.0	330	0.0
<b>Total</b>	<b>7,454,171</b>	<b>100.0</b>	<b>2,495,776</b>	<b>100.0</b>	<b>786,273</b>	<b>100.0</b>	<b>2,085,442</b>	<b>100.0</b>	<b>1,304,804</b>	<b>100.0</b>	<b>781,876</b>	<b>100.0</b>

Source: Unduplication Analysis Files

Of the persons identified as potential duplicates, 50.0 percent were male and 49.3 percent were female. It is interesting that potential duplicates that are female are identified more often within block and surrounding block, while potential duplicates that are male are identified more often within county, within state, and across state.

Table 36 presents one column with the distribution of sex for the 7,454,171 million duplicates found in the census (also shown in Table 35) and adds a second column for the 308,745,538 person records enumerated in the final 2010 Census records. The data on the duplicates are taken directly from the questionnaire without imputation or editing while the final census results have been cleaned, so there are no missing or inconsistent data for that column.

**Table 36. Sex of Duplicates Compared to All Enumerated Persons**

<b>Sex</b>	<b>Duplicates Found in the Census</b>		<b>All Persons Enumerated in the Census</b>	
	<b>Number</b>	<b>Percent</b>	<b>Number</b>	<b>Percent</b>
Male	3,725,233	50.0	151,781,326	49.2
Female	3,674,393	49.3	156,964,212	50.8
Both	90	0.0	NA	NA
Inconsistent	50,269	0.7	NA	NA
Missing	4,186	0.1	NA	NA
<b>Total</b>	<b>7,454,171</b>	<b>100.0</b>	<b>308,745,538</b>	<b>100.0</b>

Source: Unduplication Analysis Files

Using different classification systems as described before Table 36, 49.2 percent of the country was identified as male in the 2010 Census, compared to 50.0 percent of duplicates identifying as male.

## 5.2 Sampled Cases for CFU

As stated in the background, during the mid-decade testing, the unduplication cases were not successfully resolved at a high rate during followup telephone interviews. For instance, in the 2006 CFU operation, the interview was unable to confirm the person as being a duplicate in over 90 percent of cases. In the 2008 CFU operation, almost 80 percent of the potential duplicates were not resolved during CFU. These results did not support the cost it would entail to send all duplicates identified in the 2010 Census to followup. Thus, a sample of duplicate links was sent to CFU, for evaluation purposes only. These cases were sent to further assess the utility of CFU, notably in a full decennial census environment where across-state duplicates were identified (such cases are rarely found in mid-decade census tests). This section will discuss what happened in CFU to the cases that were sampled and interviewed. Note that this analysis includes cases that were considered both production and evaluation. Production cases ultimately affected Census counts, but evaluation cases did not. There is no distinction made between the two in these analyses because the goal was to determine how effective the current CFU interview would perform with duplication cases in general.

### 5.2.1 Sampling Overview

There were three broad reasons why a link could have been in the universe for sampling for CFU.

1. Experimental forms - There were two experimental forms used in the 2010 Census that pertained to coverage in the census. One of these experimental forms was mailed to a sample of households and is known as the X13 form. The other experimental form pertaining to coverage issues was used in GQs.<sup>33</sup> These two questionnaires tested experimental ways to structure questionnaires and handle responses for consideration in planning of the 2020 Census, so information obtained from a followup interview was essential. Links were flagged if either response was from one of these two experimental forms.
2. Overcount cases – As discussed in Section 5.1.11, respondents could mark on the form that someone lived or stayed at a different location occasionally.
3. Unduplication only cases – These cases did not meet the criteria to be included in either of the first two categories, and were considered within-county links, within-state links, or across-state links. Returns associated with these links could have other coverage issues such as a count discrepancy or have marked an undercount question, which are considered production coverage issues and would have been sent to CFU for resolution, but for the purpose of this analysis we consider them unduplication only.

Within these three reasons, cases were also delineated based on whether it was a HU-HU link or a HU-GQ link. In all, for sampling purposes, there were 16 strata to describe why a link was eligible for CFU. Those strata were ordered and each response link was placed into the highest order of the hierarchical list if either response in the link met the criteria for that stratum.

---

<sup>33</sup> Results from these experimental forms will be reported in the Avoid Followup Evaluation (forthcoming).

The sixteen strata are defined in order as follows:

*Experimental Forms*

1. Coverage Experiment – At least one response was from a D-1 (X13) experimental form.
2. GQ Experiment – The GQ response was from a D-20 (X1) form.

*Overcount Cases*

3. Overcount College – At least one response was identified as an overcount college response.
4. Overcount Nursing Home – At least one response was identified as an overcount nursing home response.
5. Overcount Jail – At least one response was identified as an overcount jail response.
6. Overcount Military – At least one response was identified as an overcount military response.
7. Overcount “Above the Line” Other – At least one response was identified as an overcount household multiple, person multiple, or nursing home/jail TQA response.
8. Overcount Seasonal – At least one response was identified as an overcount seasonal response.
9. Overcount Custody – At least one response was identified as an overcount custody response.
10. Overcount “Below the Line” Other – At least one response was identified as an overcount ‘for another reason’ or ‘yes only’ response.

*Unduplication Only*

11. Within County HU-HU – Not an experimental or overcount response link.
12. Within County HU-GQ – Not an experimental or overcount response link.
13. Within State HU-HU – Not an experimental or overcount response link.
14. Within State HU-GQ – Not an experimental or overcount response link.
15. Across State HU-HU – Not an experimental or overcount response link.
16. Across State HU-GQ – Not an experimental or overcount response link.

Production and evaluation CFU cases were discussed in Section 2.8, when presenting the background on CFU and the universe of cases eligible for CFU. Production cases are also referred to as “above the line” cases since they made the cut to be included in CFU, while evaluation cases are also referred to as “below the line” cases. For later reference in this report, strata 3 through 7 are considered CFU production overcount cases while strata 8 through 10 are evaluation overcount cases. Strata 1 and 2 as well as strata 11 through 16 are also considered evaluation cases.

Sampling was done within these strata, at the living quarters link level. If a HU-HU link was sampled, then everyone in that HU-HU link was eligible for evaluation if they were identified as a potential duplicate.

Table 37 below shows how the universe of duplicate links that was discussed in Section 5.1 was pared down for the desired CFU workload.

**Table 37. CFU Eligible and Sent Workload**

<b>Universe</b>	<b>Number</b>	<b>Percent of previous row</b>
Total Universe of Response Links	3,137,840	
Eligible Universe of Links	1,767,030	56.3
Links Sent to CFU	469,768	26.6

Source: Unduplication Analysis File, Weighting Spreadsheet, & Unduplication CFU Analysis Files

Section 5.1.1 reported that a total of 3,137,840 response links were found within county, within state and across state at the end of processing. Within block and surrounding block links were not eligible for CFU. Of the 3,137,840 response links, 1,767,030, or 56.3 percent, were identified as being eligible for CFU at the time of sampling. A link was considered eligible if it had been identified prior to or at the time of sample selection. Of those, 469,768 or 26.6 percent were actually selected or sampled and sent to CFU.

Table 38 shows how many of the 469,768 response links sent to CFU came from HU-HU links compared to HU-GQ links.

**Table 38. Response Links Sent to CFU, by Type of Link (unweighted)**

<b>Type of Link</b>	<b>Universe of Response Links</b>		<b>Links Sent to CFU</b>	
	<b>Number</b>	<b>Percent</b>	<b>Number</b>	<b>Percent</b>
HU-HU	2,390,271	76.2	257,079	54.7
HU-GQ	747,569	23.8	212,689	45.3
<b>Total</b>	<b>3,137,840</b>	<b>100.0</b>	<b>469,768</b>	<b>100.0</b>

Source: Unduplication CFU Analysis Files

Table 38 shows that 54.7 percent of response links sent to CFU came from HU-HU links, while 45.3 percent of response links sent to CFU came from HU-GQ links. From the total universe of response links, about 76.2 percent of response links come from HU-HU links, while 23.8 percent come from HU-GQ links. This contrast is not surprising, as Table 39 will show there were a number of HU-GQ response link categories where nearly everything that was eligible was also sent.

Table 39 shows the eligible universe of response links, the number of response links sent to CFU, and the percent of those response links in CFU compared to the eligible universe. Table 39 also shows the sample stratum that a response link was placed in, further divided by HU-HU and HU-GQ links.

**Table 39: Percent of CFU-Eligible Response Links Sent To CFU (unweighted)**

<b>Strata</b>	<b>Eligible Universe</b>	<b>Number in CFU</b>	<b>Percent in CFU</b>
Coverage Experiment HU-HU	830	786	94.7
Coverage Experiment HU-GQ	234	213	91.0
GQ Experiment	10,451	8,844	84.6
OC - College HU-HU	105,468	33,100	31.4
OC - College HU-GQ	103,037	103,012	100.0
OC - Nursing Home HU-HU	3,116	3,020	96.9
OC - Nursing Home HU-GQ	12,613	12,611	100.0
OC – Jail HU-HU	2,383	2,263	95.0
OC – Jail HU-GQ	6,984	6,977	100.0
OC - Military HU-HU	7,272	7,136	98.1
OC - Military HU-GQ	2,583	2,582	100.0
OC - Other Above HU-HU	146,500	40,104	27.4
OC - Other Above HU-GQ	19,357	19,344	99.9
OC - Seasonal HU-HU	369,323	19,147	5.2
OC - Seasonal HU-GQ	4,007	1,064	26.6
OC - Custody HU-HU	233,479	15,840	6.8
OC - Custody HU-GQ	1,098	321	29.2
OC - Other Below HU-HU	173,781	17,018	9.8
OC - Other Below HU-GQ	13,365	3,853	28.9
Within County HU-HU	192,393	42,445	22.1
Within County HU-GQ	113,874	22,803	20.0
Within State HU-HU	88,504	38,469	43.5
Within State HU-GQ	62,267	20,443	32.9
Across State HU-HU	73,278	37,751	51.5
Across State HU-GQ	20,833	10,622	51.0
<b>Total</b>	<b>1,767,030</b>	<b>469,768</b>	<b>26.6</b>

Source: Unduplication CFU Analysis Files

Table 39 indicates that there were several strata where we sent nearly all response links that were eligible. The overcount college, nursing, jail, military, and other above categories for HU-GQ links show that nearly 100 percent of all eligible links were sent to CFU. Since these categories were all production CFU categories, most of the housing unit responses for these links would have been sent to CFU unless they failed a CFU requirement, such as if another response at the MAFID had already been sent to CFU. As seen in Table 38, this is likely the reason there was such a large HU-GQ workload in CFU compared to the overall universe of response links.

Other strata with above 90 percent of the eligible response links sent include the coverage experiment stratum, as well as HU-HU links with overcount marks for nursing home, jail and military.

Overcount college HU-HU and above the line HU-HU links were only sent about 31.4 percent and 27.4 percent of the time, respectively. In these cases, using the example of an overcount college HU-HU response, there was likely a discrepancy between the HU forms, in which at least one return marked the overcount college response, but the person was duplicated at a HU instead of a GQ. It is likely the other response showed no other indication of a coverage problem, and thus never would have been sent to CFU.

All other strata generally had less than 51 percent of response links sent to CFU.

Again, there were 469,768 response links where at least one side of the link was called in CFU. The next section will show how often an interview was completed with these links.

## **5.2.2 Case Completion**

The following section will discuss how many of the 469,768 linked duplication cases that were sent to CFU had at least one response complete a CFU interview. The results on case completion will first be presented as unweighted before the weighted numbers are shown. Weighted numbers will be used for the subsequent analysis.

### **5.2.2.1 Unweighted Numbers**

Not all of the 469,768 response links sent to CFU completed a CFU interview. We completed a CFU interview with at least one housing unit in 350,757 response links, or 74.7 percent of the time.

For the purposes of classifying a link as complete, we required that at least one response from the link complete a CFU interview. For HU-GQ links, this simply means that we must talk to the HU, because CFU did not interview GQs. For HU-HU response links, it would be ideal to talk to both HUs to resolve the duplication, but this is not always possible. For example, if a person has a second home, during the course of the interview period we may only reach that person at one HU, or if a person has moved one phone number may be disconnected or reassigned. Thus having at least one completed response allows us some information to be able to evaluate whether we can resolve duplication, but may not provide the complete picture.

Table 40 shows the completion rate for response links by strata for HU to HU response links.

**Table 40: Completion Rate by Strata, HU to HU Response Link (unweighted)**

<b>Strata</b>	<b>Number in CFU</b>	<b>Number of CFU Completes</b>	<b>Completion Rate*</b>
Coverage Experiment HU-HU	786	661	84.1
OC - College HU-HU	33,100	29,483	89.1
OC - Nursing Home HU-HU	3,020	2,025	67.1
OC - Jail HU-HU	2,263	1,809	79.9
OC - Military HU-HU	7,136	6,124	85.8
OC - Other Above HU-HU	40,104	34,362	85.7
OC - Seasonal HU-HU	19,147	16,084	84.0
OC - Custody HU-HU	15,840	13,655	86.2
OC - Other Below HU-HU	17,018	14,218	83.5
Within County HU-HU	42,445	31,543	74.3
Within State HU-HU	38,469	30,230	78.6
Across State HU-HU	37,751	31,119	82.4
<b>Total</b>	<b>257,079</b>	<b>211,313</b>	<b>82.2</b>

Source: Unduplication CFU Analysis Files

\*At least one side complete

Table 40 shows that overall, 82.2 percent of HU to HU response links had at least one side complete a CFU interview. Looking across all strata, we see that the overcount nursing home and jail strata, and the unduplication-only within county and within state strata are the only ones that fall below the overall completion rate.

Table 41 shows the completion rate for response links by strata for HU to GQ response links.



**Table 41: Completion Rate by Strata, HU to GQ Response Link (unweighted)**

Strata	Number in	Number of	Completion Rate*
	CFU	CFU Completes	
Coverage Experiment HU-GQ	213	148	69.5
GQ Experiment	8,844	6,360	71.9
OC - College HU-GQ	103,012	75,621	73.4
OC - Nursing Home HU-GQ	12,611	6,147	48.7
OC - Jail HU-GQ	6,977	4,061	58.2
OC - Military HU-GQ	2,582	1,662	64.4
OC - Other Above HU-GQ	19,344	12,847	66.4
OC - Seasonal HU-GQ	1,064	614	57.7
OC - Custody HU-GQ	321	198	61.7
OC - Other Below HU-GQ	3,853	2,381	61.7
Within County HU-GQ	22,803	11,961	52.4
Within State HU-GQ	20,443	11,089	54.2
Across State HU-GQ	10,622	6,355	59.8
<b>Total</b>	<b>212,689</b>	<b>139,444</b>	<b>65.6</b>

Source: Unduplication CFU Analysis Files

\*Completed interview with HU response. No interviews were attempted with GQ response.

Table 41 shows that overall, 65.6 percent of HU to GQ response links completed a CFU interview. This rate is lower than the HU to HU response links, which may be attributable to the fact that we have an additional chance to contact a response link for HU to HU. We see that overcount college strata had the highest completion rate (73.4 percent) and that overcount nursing home had the lowest completion rate (48.7 percent).

For the remainder of this evaluation, results are presented at the person level, since that is how we will determine if we are resolving person duplication. Table 42 expands the 469,768 response links into the 565,058 person links sent to CFU. It also further divides them into HU-HU links and HU-GQ links.

**Table 42: Completed Interviews by Type of Person Link (unweighted)**

CFU completion rate	HU-HU links		HU-GQ links		Total links sent	
	Number	Percent	Number	Percent	Number	Percent
Both sides completed interview	131,199	37.2	---	---	131,199	23.2
One side completed interview	154,163	43.8	139,444	65.6	293,607	52.0
Neither side completed interview	67,007	19.0	73,245	34.4	140,252	24.8
<b>Total sent to CFU</b>	<b>352,369</b>	<b>100.0</b>	<b>212,689</b>	<b>100.0</b>	<b>565,058</b>	<b>100.0</b>

Source: Unduplication CFU Analysis Files

Of the HU-HU person links sent to CFU, interviews were conducted with both sides 37.2 percent of the time. One side was interviewed in 43.8 percent of the links. For HU-GQ links, the housing unit side was interviewed 65.6 percent of the time. Over one-third (34.4 percent) of all

HU-GQ links sent to CFU were not contacted while 19.0 percent of HU-HU links sent to CFU were not contacted.

Table 43 below shows whether a HU-HU person link completed an interview at both housing units, or just at one housing unit, by stratum.

HU-HU response links have at least one person link across the responses, but can have more than that. For that reason, the number of person links will be larger than the number of response links. HU-GQ response links are always a one-to-one match, and thus will be equal across response and person links.

**Table 43: Completed Interviews by Strata for HU-HU Person Links (unweighted)**

Strata	Both sides completed		One side completed		Total Number of cases
	Number	Percent	Number	Percent	
Coverage Experiment HU-HU	429	40.2	460	43.1	<b>1,067</b>
OC - College HU-HU	17,443	47.5	14,963	40.7	<b>36,739</b>
OC - Nursing Home HU-HU	727	22.5	1,458	45.1	<b>3,234</b>
OC - Jail HU-HU	872	30.6	1,378	48.4	<b>2,486</b>
OC - Military HU-HU	3,850	41.6	4,014	43.4	<b>9,253</b>
OC - Other Above HU-HU	26,882	42.3	27,308	42.9	<b>63,579</b>
OC - Seasonal HU-HU	12,895	43.2	12,013	40.2	<b>29,837</b>
OC - Custody HU-HU	9,599	40.6	10,824	45.7	<b>23,670</b>
OC - Other Below HU-HU	8,380	38.4	9,609	44.1	<b>21,771</b>
Within County HU-HU	18,274	28.1	28,954	44.6	<b>64,908</b>
Within State HU-HU	15,931	31.6	22,900	45.4	<b>50,479</b>
Across State HU-HU	15,917	35.4	20,282	45.1	<b>44,987</b>
<b>Total</b>	<b>131,199</b>	<b>37.2</b>	<b>154,163</b>	<b>43.8</b>	<b>352,369</b>

Source: Unduplication CFU Analysis Files

Table 43 again shows that overall, 37.2 percent of all person links had CFU interviews completed at both HUs, while 43.8 percent only had interviews completed at one HU. We notice that most strata were generally close or better than the overall average at completing interviews with both sides. However, unduplication-only person links show that interviews were completed at both housing units less than the average. We also notice that overcount nursing home and jail links were completed at both housing units considerably less often than in other strata. This may be because they are really in a GQ listed as a HU, and we will never be able to contact them at both. However, those strata appear to have been better at completing an interview with at least one of the HUs.

See Table 41 for similar HU to GQ results. Housing Unit to GQ links are one-to-one links, so there would be no difference.

### 5.2.2.2 Weighted Numbers

In the previous section, we showed the completion rates using unweighted numbers. Since response links were sampled and weights were applied, Table 44 shows how the person links are weighted up to be representative of the unduplication type they represent (for example, overcount college). These weighted numbers will be used throughout the remainder of the CFU analysis in this evaluation.

Table 44 compares the unweighted to weighted person numbers when a CFU interview was completed, since only persons from completed interview links were used in the analysis. Table 44 is presented by stratum, but the coverage experiment strata and the GQ experiment stratum are no longer present. For evaluation purposes, we decided to place each response link from those strata into the strata that they would have been placed into ignoring their experimental status. This allows for easier comparison of cases that had marked the overcount box compared with cases that had not. Full results of experimental cases can be found in the Avoid Followup Evaluation.

**Table 44: Number of Unweighted and Weighted Person Links From Completed Interviews**

<b>Strata</b>	<b>Unweighted</b>	<b>Weighted</b>
OC - College HU-HU	32,530	109,894
OC - College HU-GQ	80,878	81,367
OC - Nursing Home HU-HU	2,185	3,146
OC - Nursing Home HU-GQ	6,235	6,381
OC – Jail HU-HU	2,251	2,804
OC – Jail HU-GQ	4,148	4,245
OC - Military HU-HU	7,882	8,904
OC - Military HU-GQ	1,667	1,695
OC - Other Above HU-HU	54,314	202,805
OC - Other Above HU-GQ	13,570	13,640
OC - Seasonal HU-HU	25,195	563,492
OC - Seasonal HU-GQ	628	3,652
OC - Custody HU-HU	20,493	325,343
OC - Custody HU-GQ	204	1,016
OC - Other Below HU-HU	18,096	203,296
OC - Other Below HU-GQ	2,453	12,395
Within County HU-HU	47,314	281,046
Within County HU-GQ	12,068	108,473
Within State HU-HU	38,871	112,237
Within State HU-GQ	11,194	59,200
Across State HU-HU	36,231	84,381
Across State HU-GQ	6,399	20,151
<b>Total</b>	<b>424,806</b>	<b>2,209,562</b>

Source: Unduplication CFU Analysis Files

Table 44 shows that 424,806 person links had at least one completed CFU interview, which when weighted represents 2,209,562 persons links. Certain strata such as HU-GQ links for overcount college, nursing home, jail, military, and other “above the line” cases are not heavily weighted, while other strata are, such as all six of the unduplication-only strata. This relates back to Table 39 and how much of the eligible universe was represented by the cases sent to CFU.

Since a CFU interview could have been completed with both sides of a HU-HU link (but with only the housing unit side of a HU-GQ link), Table 45 shows how often an interview was completed with one or both sides of a person link. There are no numbers for ‘Neither side completed interview’ because person links were only weighted when at least one side completed an interview.

**Table 45: Completion Rates by Type of Person Link (Weighted)**

CFU Completion Rate	HU-HU links		HU-GQ links		Total links sent	
	Number	Percent (Std. Error)	Number	Percent	Number	Percent (Std. Error)
Both sides completed interview	899,888	47.4 (0.2)	---	---	899,888	40.7 (0.2)
One side completed interview	997,459	52.6 (0.2)	312,215	100.0	1,309,674	59.3 (0.2)
<b>Total sent to CFU</b>	<b>1,897,347</b>	<b>100.0</b>	<b>312,215</b>	<b>100.0</b>	<b>2,209,562</b>	<b>100.0</b>

Source: Unduplication CFU Analysis Files

Similar to Table 42, Table 45 shows with weighted numbers that we more often completed an interview with only one side of a person link instead of with both sides. For HU-HU links, 52.6 percent only completed one CFU interview while 47.4 percent of the links completed an interview with each side.

The following table builds from the HU-HU links in Table 45. Table 46 shows how many person links in each stratum completed interviews with both sides instead of just one side.

**Table 46: Completed Person Links by Stratum, for HU-HU links (Weighted)**

Strata	Completed Interview with Both Sides		Completed Interview with One Side		Row Total
	Number	Percent (Std. Error)	Number	Percent (Std. Error)	Number
OC - College HU-HU	58,442	53.18 (0.02)	51,451	46.82 (0.02)	109,894
OC - Nursing Home HU-HU	1,037	32.0 (1.7)	2,108	67.0 (1.7)	3,146
OC - Jail HU-HU	1,079	38.5 (0.3)	1,724	61.5 (0.3)	2,804
OC - Military HU-HU	4,343	48.8 (0.5)	4,561	51.2 (0.5)	8,904
OC - Other Above HU-HU	96,943	47.8 (0.3)	105,863	52.2 (0.3)	202,805
OC - Seasonal HU-HU	297,720	52.8 (0.5)	265,772	47.2 (0.5)	563,492
OC - Custody HU-HU	151,990	46.7 (0.5)	173,354	53.3 (0.6)	325,343
OC - Other Below HU-HU	96,495	47.5 (0.2)	106,801	52.5 (0.2)	203,296
Within County HU-HU	108,716	38.7 (0.1)	172,329	61.3 (0.1)	281,046
Within State HU-HU	46,032	41.0 (0.4)	66,206	59.0 (0.4)	112,237
Across State HU-HU	37,091	43.0 (0.1)	47,290	56.0 (0.1)	84,381
<b>Total</b>	<b>899,888</b>	<b>47.4 (0.2)</b>	<b>997,459</b>	<b>52.6 (0.2)</b>	<b>1,897,347</b>

Source: Unduplication CFU Analysis Files

The person links between two housing units where one marked a nursing home overcount category were the least likely to complete interviews with both sides. This could partially reflect a misclassification or another issue, as nursing homes (and jails, the second least likely to have both sides complete interviews) are not expected to be housing unit to housing unit duplication.

The number of complete interviews by stratum for HU-GQ cases was contained in Table 44. Since only one side could be contacted, there is no corresponding table to Table 46 for HU-GQ links.

Completion percentage for one particular subset of interest is the HU-HU links that had the same phone number on each side of the link. These cases would have received two distinct phone calls during the CFU operation, and been asked the same series of questions in each interview. Table 47 shows how often we successfully completed an interview with both sides of these links.

**Table 47: Completion Percentage from HU-HU Person Links, by Phone Number Match**

CFU completion percentage	HU-HU links with the same phone number		HU-HU links with different phone numbers		Total HU-HU links sent	
	Number	Percent (Std. Error)	Number	Percent (Std. Error)	Number	Percent (Std. Error)
Both sides completed interview	177,275	69.0 (0.3)	722,613	44.0 (0.2)	899,888	47.4 (0.2)
One side completed interview	79,481	31.0 (0.3)	917,978	56.0 (0.2)	997,459	52.6 (0.2)
<b>Total sent to CFU</b>	<b>256,756</b>	<b>100.0</b>	<b>1,640,591</b>	<b>100.0</b>	<b>1,897,347</b>	<b>100.0</b>

Source: Unduplication CFU Analysis Files

Table 47 shows that when two HU responses shared the same phone number, both sides completed an interview in CFU 69.0 percent of the time, but only one interview was completed in 31.0 percent of cases. An alternative approach to these cases was tried in the qualitative interviews and will be discussed in Section 5.4.

Sampling for unduplication responses for CFU had to occur before the entire universe of duplicate links had been identified. It is useful to note how the universe at the time of selection (the CFU universe) compares to the final overall universe. Table 48 uses the type of forms linked (notably mail returns or enumerator-completed returns) to compare the CFU universe to the entire universe that was described in Section 5.1. The overall universe of person links in the right column of Table 48 is from Table 8 and is subset to contain only the person links that were found within county, within state, and across state lines. Links within block or within surrounding blocks of each other are omitted from this table since they were not eligible for CFU.

**Table 48: Distribution of Linked Form Types, by CFU Universe and Overall**

	CFU Universe (Unweighted)		CFU Universe (Weighted)		Overall Universe	
	Number	Percent	Number	Percent (Std. Error)	Number	Percent
Mail to Enum	46,782	11.01	272,870	12.35 (0.08)	1,556,364	37.30
Mail to Mail	236,571	55.69	1,613,038	73.00 (0.10)	1,672,543	40.09
Enum to Enum	2,009	0.47	11,439	0.52 (0.02)	195,646	4.69
Mail to GQ	136,200	32.06	291,391	13.19 (0.03)	616,746	14.78
Enum to GQ	3,244	0.76	20,824	0.94 (0.02)	130,823	3.14
<b>Total</b>	<b>424,806</b>	<b>100.00</b>	<b>2,209,562</b>	<b>100.00</b>	<b>4,172,122</b>	<b>100.00</b>

Source: Unduplication CFU Analysis Files

From Table 48, the distribution of form type for the CFU universe of person links (using either the column with the weighted or unweighted percent) looks vastly different from the form types seen in the overall universe of person links. This can be attributed to a large number of enumerator (and GQ) responses that were not data captured until after the CFU sample selection had occurred. For more detailed information about the availability of person records, please refer

to Appendix B. Thus, any conclusion made about the CFU cases in this evaluation may not be true for the entire population of persons potentially identified as duplicates. The weighted number of person links from one mailback form to another mailback form (1,612,038) was fairly close to the overall number of mailback to mailback person links (1,672,543). However, in the weighted CFU universe, this type of linkage was vastly overrepresented, accounting for nearly 73 percent of the universe, when it was only 40.09 percent of the overall universe. The person links with an enumerator return accounting for at least one of the linked returns were underrepresented in the CFU universe, compared to the overall universe.

### **5.2.3 Complex Situations**

The CFU interview contained probes to identify people who were not initially included on the household roster as well as people who were on the roster but (according to the census residence rule) should not have been enumerated at the housing unit. Regardless of the source of coverage improvement, all households sent for followup received the same core questions to identify missed and erroneously enumerated people. Information gathered during the initial enumeration was passed to the CFU interview, and respondents added or deleted people from the roster of the initial return.

The probes that addressed other living situations were of the most interest in this evaluation. Those probes asked if anyone may have been living away from the CFU address for any of the following reasons:

- Moving around Census Day
- Attending College
- A child living part of the time with someone else
- Away for Military Service
- Away for a job
- Because they have a seasonal home or second home
- Staying somewhere else for an extended time or living part of the time at another residence
- Staying in a GQ

Since we have identified persons as potential duplicates, there is an expectation that the CFU interview should result in the mention of a complex living situation for each of these persons.

For more information on the 2010 CFU interview, please see the CFU Application Design Document (IBM, 2007).

#### **5.2.3.1 Mention of Any Complex Living Situation**

Table 49 shows the number and percent of potential duplicates for whom a complex living situation was indeed mentioned during a CFU interview. We will further investigate cases in Table 51 where both sides completed a CFU interview to see how often both sides mention a complex living situation.

**Table 49: Rate of Complex Living Situations Being Mentioned in Person Links**

<b>Mention Complex Living Situation</b>	<b>Number</b>	<b>Percent (Std. Error)</b>
At least one side mentioned	1,296,870	58.7 (0.1)
<i>Both Sides</i>	383,771	17.4 (0.1)
<i>One Side</i>	913,099	41.3 (0.1)
Neither side mentioned	912,692	41.3 (0.1)
<b>Total</b>	<b>2,209,562</b>	<b>100.0</b>

Source: Unduplication CFU Analysis Files

Table 49 indicates that overall about 58.7 percent of all person links mentioned a complex living situation. About 41.3 percent of person links resulted in only one side mentioning a complex living situation, while only about 17.4 percent of person links resulted in both sides mentioning a complex living situation.

Table 50 below shows how often a complex living situation was mentioned by at least one side by stratum. This is one measure of the success of a CFU interview, since the first step to resolving a complex living situation is getting the respondent to acknowledge or identify one.



**Table 50: Rate of Complex Living Situations Being Mentioned for Person Links, by Stratum**

Strata	Total Number	Mention Complex Living Situation	
		Number	Percent (Std. Error)
OC - College HU-HU	109,894	94,096	85.6 (0.6)
OC - College HU-GQ	81,367	78,331	96.3 (0.1)
OC - Nursing Home HU-HU	3,146	1,961	62.3 (0.2)
OC - Nursing Home HU-GQ	6,381	5,295	83.0 (0.3)
OC - Jail HU-HU	2,804	789	28.1 (0.2)
OC - Jail HU-GQ	4,245	3,282	77.3 (0.4)
OC - Military HU-HU	8,904	4,101	46.1 (0.5)
OC - Military HU-GQ	1,695	1,241	73.2 (0.1)
OC - Other Above HU-HU	202,805	135,178	66.7 (0.3)
OC - Other Above HU-GQ	13,640	11,796	86.5 (0.3)
OC - Seasonal HU-HU	563,492	428,583	76.1 (0.2)
OC - Seasonal HU-GQ	3,652	1,849	50.6 (5.0)
OC - Custody HU-HU	325,343	256,219	78.8 (0.4)
OC - Custody HU-GQ	1,016	573	56.4 (14.3)
OC - Other Below HU-HU	203,296	89,269	43.9 (0.2)
OC - Other Below HU-GQ	12,395	5,642	45.5 (0.9)
Within County HU-HU	281,046	56,895	20.2 (0.2)
Within County HU-GQ	108,473	42,037	38.8 (0.2)
Within State HU-HU	112,237	28,439	25.3 (0.1)
Within State HU-GQ	59,200	29,261	49.4 (0.2)
Across State HU-HU	84,381	13,992	16.6 (0.3)
Across State HU-GQ	20,151	8,044	39.9 (1.0)
<b>Total</b>	<b>2,209,562</b>	<b>1,296,870</b>	<b>58.7 (0.1)</b>

Source: Unduplication CFU Analysis Files

Of the six unduplication-only strata, none resulted in a complex living situation being mentioned more than half the time. Within-state HU-GQ links were the most successful, with 49.4 percent mentioning a complex living situation. Only four of the sixteen overcount strata were below a fifty percent success rate in Table 50. Two of those four were from the ‘other below’ strata, which means they had either marked the ‘for another reason’ overcount box or only marked a ‘yes’ response without choosing a category on the initial return. These cases would be less likely to see their situation in any of the CFU probes. Generally, the overcount strata mention a complex situation more often than the unduplication-only categories. This makes sense, since at least one of the two linked responses from the overcount strata had already indicated that a complex living situation existed on the original response, whereas respondents on the unduplication-only cases did not indicate a situation or did not know about a situation.

Table 51 shows the number and percent of potential duplicates where a complex living situation was mentioned during a CFU interview by stratum and both HUs completed a CFU interview.

**Table 51: Rate of Complex Living Situations Being Mentioned for HU-HU Person Links When Both Sides Completed a CFU Interview, by Stratum**

Strata	Both sides completed in CFU						Row Total
	Both sides mentioned a complex Living Situation		One side mentioned a complex Living Situation		Neither side mentioned a complex Living Situation		
	Number	Percent (Std. Error)	Number	Percent (Std. Error)	Number	Percent (Std. Error)	
OC - College HU-HU	45,084	77.1 (0.6)	7,934	13.6 (0.1)	5,425	9.3 (0.7)	58,442
OC - Nursing Home HU-HU	310	29.9 (0.2)	421	40.6 (2.7)	307	29.6 (2.5)	1,037
OC – Jail HU-HU	94	8.7 (0.3)	332	30.8 (0.1)	654	60.6 (0.4)	1,079
OC – Military HU-HU	844	19.44 (0.03)	1,665	38.3 (0.1)	1,833	42.2 (0.1)	4,343
OC - Other Above HU-HU	42,278	43.6 (1.0)	33,834	34.9 (0.5)	20,831	21.5 (0.5)	96,943
OC - Seasonal HU-HU	174,759	58.7 (0.6)	81,199	27.3 (0.5)	41,763	14.0 (0.1)	297,720
OC - Custody HU-HU	83,832	55.2 (0.1)	52,611	34.6 (0.6)	15,547	10.2 (0.7)	151,990
OC - Other Below HU-HU	21,294	22.1 (0.1)	34,079	35.3 (0.2)	41,121	42.6 (0.3)	96,495
Within County HU-HU	7,605	7.0 (0.1)	21,755	20.0 (0.1)	79,356	73.0 (0.2)	108,716
Within State HU-HU	5,598	12.2 (0.1)	9,343	20.30 (0.01)	31,091	67.54 (0.05)	46,032
Across State HU-HU	2,073	5.6 (0.1)	5,382	14.5 (0.1)	29,637	79.9 (0.3)	37,091
Total	383,771	42.6 (0.2)	248,554	27.6 (0.2)	267,563	29.7 (0.1)	899,888

Source: Unduplication CFU Analysis Files

When CFU interviews were completed with both sides of a person link, 42.6 percent of the time a complex living situation was mentioned in both interviews. Neither side mentioned a complex living situation 29.7 percent of the time however. Overcount strata that had marked the college, seasonal, and custody boxes initially were the most likely to mention a complex living situation in both interviews.

Table 52 shows the number and percent of potential duplicates where a complex living situation was mentioned during a CFU interview by stratum and only one HU completed a CFU interview.

**Table 52: Rate of Complex Living Situations Being Mentioned for HU-HU Person Links When One Side Completes a CFU Interview, by Stratum**

Strata	One side completed in CFU				Row Total
	One Side Mentioned a complex Living Situation		Neither Side Mentioned a complex Living Situation		
	Percent		Percent		
	Number	(Std. Error)	Number	(Std. Error)	
OC - College HU-HU	41,078	79.8 (0.6)	10,373	20.2 (0.6)	51,451
OC - Nursing Home HU-HU	1,230	58.3 (1.8)	878	41.7 (1.8)	2,108
OC – Jail HU-HU	363	21.1 (0.5)	1,361	78.9 (0.5)	1,724
OC - Military HU-HU	1,591	34.9 (0.7)	2,970	65.1 (0.7)	4,561
OC - Other Above HU-HU	59,066	55.79 (0.03)	46,797	44.21 (0.03)	105,863
OC - Seasonal HU-HU	172,625	65.0 (0.3)	93,147	35.0 (0.3)	265,772
OC - Custody HU-HU	119,776	69.1 (0.4)	53,578	30.9 (0.4)	173,354
OC - Other Below HU-HU	33,895	31.7 (0.2)	72,906	68.3 (0.2)	106,801
Within County HU-HU	27,534	16.0 (0.2)	144,795	84.0 (0.2)	172,329
Within State HU-HU	13,498	20.4 (0.3)	52,708	79.6 (0.3)	66,206
Across State HU-HU	6,537	13.8 (0.4)	40,753	86.2 (0.4)	47,290
Total	477,195	47.8 (0.1)	520,264	52.2 (0.1)	997,459

Source: Unduplication CFU Analysis Files

When a CFU interview was only completed with one side, the overcount strata for college, nursing home, seasonal and custody were the most likely to mention a complex living situation.

Table 53 shows the number and percent of HU-GQ person links where a person mentioned a complex situation by strata. Since only the HU side was eligible to be called in CFU, there is only one chance for a complex living situation to be mentioned.

**Table 53: Rate of Complex Living Situations Being Mentioned for HU-GQ Person Links, by Stratum**

<b>Strata</b>	<b>One Side Mentioned</b>		<b>Neither Side</b>		<b>Row Total</b>
	<b>Number</b>	<b>Percent (Std. Error)</b>	<b>Number</b>	<b>Percent (Std. Error)</b>	<b>Number</b>
OC - College HU-GQ	78,331	96.3 (0.1)	3,036	3.7 (0.1)	81,367
OC - Nursing Home HU-GQ	5,295	83.0 (0.3)	1,086	17.0 (0.3)	6,381
OC – Jail HU-GQ	3,282	77.3 (0.4)	963	22.7 (0.4)	4,245
OC - Military HU-GQ	1,241	73.2 (0.1)	454	26.8 (0.1)	1,695
OC - Other Above HU-GQ	11,796	86.5 (0.3)	1,845	13.5 (0.3)	13,640
OC - Seasonal HU-GQ	1,849	50.6 (5.0)	1,803	49.4 (5.0)	3,652
OC - Custody HU-GQ	573	56.4 (14.3)	443	43.6 (14.3)	1,016
OC - Other Below HU-GQ	5,642	45.5 (0.9)	6,752	54.5 (0.9)	12,395
Within County HU-GQ	42,037	38.8 (0.2)	66,436	61.2 (0.2)	108,473
Within State HU-GQ	29,261	49.4 (0.2)	29,940	50.6 (0.2)	59,200
Across State HU-GQ	8,044	39.9 (1.0)	12,107	60.1 (1.0)	20,151
<b>Total</b>	<b>187,350</b>	<b>60.0 (0.2)</b>	<b>124,865</b>	<b>40.0 (0.2)</b>	<b>312,215</b>

Source: Unduplication CFU Analysis Files

Table 53 indicates that overcount college HU-GQ links mentioned a complex living situation the most at 96.3 percent, while within-county HU-GQ links mentioned a complex living situation the least at 38.8 percent of the time. The production overcount strata all had a success rate of 73.2 percent or higher, which is a good indication that persons recognized these complex living situations. The non-production overcount reasons did not fare quite as well. However, seasonal and custody cases are not expected to be HU-GQ duplication, so those links could reflect a misclassification or other issue. Similar to previous results, the unduplication-only strata did not perform as well as the overcount strata. However, they are performing better than their HU-HU counterparts.

#### 5.2.3.2 Mention of Specific Complex Living Situation, Overcount Strata

For cases from the overcount strata, we have an expectation based on the overcount marking from the initial census return of what their complex living situation is. In the CFU interview, we would expect the respondent to describe the same complex living situation. Table 54 shows how often the CFU response aligned with the initial census response in this regard. Cases are classified as follows, using an example of a case that had marked the college overcount box:

- Yes – The CFU response(s) mentioned a college situation and only a college situation.
- No – The CFU response(s) mentioned a complex living situation, but it did not align with the overcount reason (which in this example is “college”).
- Multiple Living Situations Mentioned for One Person or Different Living Situation Mentioned Across Both Sides – The CFU response(s) mentioned multiple complex living situations, or each side mentioned a different complex living situation. For example,
  - Mentioning both “child custody” and “a seasonal or second home”, or
  - One side indicated “College” and the other side indicated “Military”.
- No Mention – Neither side mentioned a complex living situation.

**Table 54: Frequency that the CFU Complex Situation Mentioned Agreed with the Overcount Strata**

Strata	Does the Complex Situation Meet Expectations?								Row Total
	Yes		No		Multiple or Different		No Mention		
	Number	Percent	Number	Percent	Number	Percent	Number	Percent	
		(Std. Error)		(Std. Error)		(Std. Error)		(Std. Error)	
OC - College HU-HU	74,972	68.2 (0.1)	5,535	5.04 (0.04) 0.434	13,590	12.4 (0.5) 1.245	15,797	14.4 (0.6)	109,894
OC - College HU-GQ	76,965	94.6 (0.1)	353	(0.001)	1,013	(0.0008)	3,036	3.7 (0.1)	81,367
OC - Nursing Home HU-HU	1,124	35.7 (0.5)	390	12.4 (0.4)	447	14.2 (1.1)	1,185	37.7 (0.2)	3,146
OC - Nursing Home HU-GQ	4,322	67.7 (0.1)	312	4.9 (0.1)	661	10.4 (0.4)	1,086	17.0 (0.3)	6,381
OC – Jail HU-HU	130	4.6 (0.1)	589	21.0 (0.7)	70	2.5 (0.1)	2,015	71.9 (0.2)	2,804
OC – Jail HU-GQ	2,885	68.0 (0.2)	184	4.3 (0.3)	214	5.0 (0.3)	963	22.7 (0.4)	4,245
OC - Military HU-HU	1,025	11.5 (0.4)	2,318	26.03 (0.03)	758	8.5 (0.1)	4,803	53.9 (0.5)	8,904
OC - Military HU-GQ	1,053	62.1 (0.3)	105	6.2 (0.1)	83	4.9 (0.3)	454	26.8 (0.1)	1,695
OC - Seasonal HU-HU	296,293	52.58 (0.03)	72,183	12.8 (0.1)	60,106	10.7 (0.3)	134,910	23.9 (0.2)	563,492
OC - Seasonal HU-GQ	557	15.3 (3.1)	895	24.5 (0.9)	397	10.9 (2.8)	1,803	49.4 (5.0)	3,652
OC - Custody HU-HU	235,443	72.4 (0.5)	6,392	2.0 (0.2)	14,384	4.4 (0.1)	69,125	21.2 (0.4)	325,344
OC - Custody HU-GQ	235	23.1 (6.8)	209	20.6 (1.5)	129	12.7 (6.0)	443	43.6 (14.3)	1,016
Total	695,004	62.5 (0.1)	89,465	8.0 (0.2)	91,852	8.3 (0.2)	235,619	21.2 (0.2)	1,111,939

Source: Unduplication CFU Analysis Files

College HU-GQ cases met expectations 94.6 percent of the time, followed by nursing home HU-GQ cases and jail HU-GQ cases. The cases that met expectations least often are those cases where the overcount category seems misaligned from the case designation as a HU-HU or HU-GQ link. For instance, jail HU-HU cases only met expectations 4.6 percent of the time and seasonal HU-GQ cases only met expectations 15.3 percent of the time.

The strata for ‘above the line – other’ and ‘below the line - other’ were omitted from Table 49 since they did not have one specific overcount category that could be mapped to a specific CFU probe. The “above the line-other” cases are largely person multiple and household multiple cases, while “below the line-other” cases are ‘for another reason’ or ‘yes only’ cases. Since there is not an

expectation for which complex living situation they should mention, Table 55 shows all the living situations that were mentioned for these HU-HU person links.

**Table 55: Complex Situations for Other Above the Line and Below the Line Overcount Cases – HU to HU Person Links**

<b>Complex Living Situation</b>	<b>Above The Line Other</b>		<b>Below the Line Other</b>	
	<b>Number</b>	<b>Percent (Std. Error)</b>	<b>Number</b>	<b>Percent (Std. Error)</b>
No Mention of a Complex Living Situation	67,628	33.3 (0.3)	114,027	56.1 (0.2)
Mention Complex Living Situation	135,178	66.7 (0.3)	89,269	43.9 (0.2)
<i>College</i>	17,306	12.8 (0.2)	11,710	13.1 (0.1)
<i>Military</i>	336	0.25 (0.03)	40	0.05 (0.04)
<i>Job</i>	2,944	2.18 (0.01)	6,525	7.3 (0.2)
<i>Custody</i>	56,767	41.99 (0.04)	10,790	12.1 (0.7)
<i>Seasonal</i>	28,273	20.9 (0.3)	20,820	23.3 (0.3)
<i>Another Reason</i>	6,583	4.9 (0.3)	14,115	15.8 (0.8)
<i>GQ</i>	461	0.34 (0.01)	525	0.6 (0.2)
<i>Move</i>	2,526	1.9 (0.1)	6,215	7.0 (0.4)
<i>Multiple Living Situations Mentioned</i>	7,909	5.9 (0.2)	5,487	6.1 (0.1)
<i>Different Living Situations Across Responses</i>	10,630	7.9 (0.5)	10,492	11.8 (0.3)
<i>Don't Know/Roster Duplicate</i> <sup>34</sup>	1,442	1.07 (0.01)	2,550	2.9 (0.1)
<b>Total</b>	<b>202,805</b>	<b>100.0</b>	<b>203,296</b>	<b>100.0</b>

Source: Unduplication CFU Analysis Files

For “other above the line” HU-HU person links that mentioned a complex living situation in CFU, Table 55 indicates that custody (41.99 percent), seasonal (20.9 percent) or college situations (12.8 percent) were mentioned most often. For below the line HU-HU person links, there is more of a spread of complex living situations that are mentioned. The most prominent situations are college, custody, seasonal, and another reason. Respondents also mentioned different situations between the two sides of a link 11.8 percent of the time.

<sup>34</sup> After verifying the address of the CFU household, the interviewer next verified the roster from the original enumeration. During this roster verification, a respondent indicated that they did not know someone on the roster or that the person was already listed elsewhere on the roster.

Table 56 shows the living situations that were mentioned for ‘other above the line’ and ‘other below the line’ HU-GQ person links.

**Table 56: Complex Situations for Above the Line and Below the Line Overcount Cases – HU to GQ Person Links**

<b>Complex Living Situation</b>	<b>Above The Line</b>		<b>Below the Line</b>	
	<b>Number</b>	<b>Percent (Std. Error)</b>	<b>Number</b>	<b>Percent (Std. Error)</b>
No Mention of a Complex Living Situation	1,845	13.5 (0.3)	6,752	54.5 (0.9)
Mention Complex Living Situation	11,796	86.5 (0.3)	5,642	45.5 (0.9)
<i>College</i>	8,518	72.2 (0.4)	368	6.5 (1.5)
<i>Military</i>	304	2.58 (0.01)	38	0.7 (0.6)
<i>Job</i>	36	0.3 (0.1)	212	3.7 (0.8)
<i>Custody</i>	65	0.55 (0.05)	258	4.6 (1.3)
<i>Seasonal</i>	48	0.41 (0.01)	255	4.5 (0.7)
<i>Another Reason</i>	90	0.8 (0.1)	462	8.2 (2.9)
<i>GQ</i>	1,905	16.1 (0.3)	2,456	43.5 (5.8)
<i>Move</i>	57	0.5 (0.1)	188	3.33 (0.03)
<i>Multiple Living Situations Mentioned</i>	693	5.9 (0.2)	1,260	22.3 (2.3)
<i>Don't Know/Roster Duplicate</i> <sup>35</sup>	79	0.67 (0.04)	147	2.6 (0.4)
<b>Total</b>	<b>13,640</b>	<b>100.0</b>	<b>12,395</b>	<b>100.0</b>

Source: Unduplication CFU Analysis Files

For “other above the line” HU-GQ person links, Table 56 shows that college (72.2 percent) and GQ (16.1 percent) living situations were mentioned most often. For “other below the line” HU-GQ person links, respondents most often mentioned a GQ (43.5percent) living situation or mentioned multiple complex living situations (22.3 percent). For those below the line cases that mentioned a GQ living situation, they most often mentioned “Some other type of GQ” or a Group Home.

<sup>35</sup> After verifying the address of the CFU household, the interviewer next verified the roster from the original enumeration. During this roster verification, a respondent indicated that they did not know someone on the roster or that the person was already listed elsewhere on the roster.



### 5.2.3.3 Mention of Specific Complex Living Situation, Unduplication-only Strata

Similar to the above the line and below the line overcount other cases, for unduplication only cases there is not an expectation for which complex situation should be mentioned. Table 57 shows the different living situations that were mentioned for within county, within state and across state HU-HU unduplication only links. The column totals were initially presented in Table 50.

**Table 57: Type of Complex Situations for Unduplication-Only HU-HU Links**

<b>Complex Living Situation</b>	<b>Within-County</b>		<b>Within-State</b>		<b>Across-State</b>	
	<b>Number</b>	<b>Percent (Std. Error)</b>	<b>Number</b>	<b>Percent (Std. Error)</b>	<b>Number</b>	<b>Percent (Std. Error)</b>
No Mention of a Complex Living Situation	224,150	79.8 (0.2)	83,799	74.7 (0.2)	70,389	83.4 (0.3)
Mention Complex Living Situation	56,895	20.2 (0.2)	28,438	25.3 (0.2)	13,991	16.6 (0.3)
<i>Move</i>	9,445	16.6 (0.5)	2,976	10.5 (0.5)	1,154	8.2 (0.7)
<i>College</i>	14,762	25.9 (0.3)	8,353	29.37 (0.03)	2,505	17.9 (0.6)
<i>Military</i>	75	0.13 (0.03)	25	0.09 (0.03)	38	0.3 (0.1)
<i>Job</i>	537	0.9 (0.1)	740	2.6 (0.1)	872	6.2 (0.1)
<i>Custody</i>	6,185	10.87 (0.04)	1,668	5.9 (0.1)	409	2.9 (0.2)
<i>Seasonal</i>	8,867	15.6 (0.1)	7,884	27.7 (0.7)	5,810	41.5 (1.5)
<i>Another Reason</i>	8,719	15.3 (0.2)	3,110	10.9 (0.1)	1,523	10.9 (0.4)
<i>GQ</i>	1,166	2.0 (0.3)	307	1.1 (0.1)	147	1.1 (0.1)
<i>Multiple Living Situations Mentioned</i>	2,654	4.66 (0.05)	1,128	4.0 (0.4)	623	4.5 (0.2)
<i>Different Living Situations Across Responses</i>	2,353	4.1 (0.3)	1,478	5.20 (0.04)	602	4.30 (0.04)
<i>Don't Know/Roster Duplicate</i> <sup>36</sup>	2,132	3.75 (0.04)	769	2.7 (0.2)	308	2.2 (0.2)
<b>Total</b>	<b>281,045</b>	<b>100.0</b>	<b>112,237</b>	<b>100.0</b>	<b>84,380</b>	<b>100.00</b>

Source: Unduplication CFU Analysis Files

<sup>36</sup> After verifying the address of the CFU household, the interviewer next verified the roster from the original enumeration. During this roster verification, a respondent indicated that they did not know someone on the roster or that the person was already listed elsewhere on the roster.

For within-county links that mentioned a complex living situation, respondents mentioned a college situation the most (25.9 percent of the time). Moving, seasonal or second residence, another unspecified reason, and custody were each mentioned at least ten percent of the time as well.

For within-state links that mentioned a complex living situation, respondents mentioned a college situation or a seasonal or second home situation most frequently (29.37 percent and 27.7 percent, respectively). Other living situations that were mentioned at least ten percent of the time were moving (10.5 percent) and for another unspecified reason (10.9 percent).

For across-state links, respondents mentioned a seasonal or second home more often than any other situation (41.5 percent). College and another unspecified reason were each mentioned at least ten percent of the time as well.

Across all three geographies, respondents consistently mentioned the complex living situations of college, seasonal/second home, another unspecified reason, and, to a lesser degree, moving.

Table 58 shows the number and percent of all of the different living situations that were mentioned for within county, within state and across state HU-GQ unduplication only links. For this table we also looked at the specific type of GQ mentioned.

**Table 58: Type of Complex Situations for Unduplication-Only HU-GQ Links**

<b>Complex Living Situation</b>	<b>Within-County</b>		<b>Within-State</b>		<b>Across-State</b>	
	<b>Number</b>	<b>Percent (Std. Error)</b>	<b>Number</b>	<b>Percent (Std. Error)</b>	<b>Number</b>	<b>Percent (Std. Error)</b>
No Mention of a Complex Living Situation	66,436	61.2 (0.2)	29,940	50.6 (0.2)	12,107	60.1 (1.0)
Mention Complex Living Situation	42,037	38.8 (0.2)	29,261	49.4 (0.2)	8,042	39.9 (1.0)
<i>College</i>	8,803	20.9 (0.4)	15,200	51.9 (0.2)	5,850	72.7 (1.1)
<i>Military</i>	278	0.66 (0.05)	92	0.31 (0.05)	286	3.6 (0.4)
<i>Job</i>	365	0.87 (0.05)	270	0.92 (0.05)	170	2.1 (0.2)
<i>Custody</i>	283	0.7 (0.1)	339	1.2 (0.3)	90	1.1 (0.5)
<i>Seasonal</i>	473	1.1 (0.4)	225	0.8 (0.1)	158	2.0 (0.6)
<i>Another Reason</i>	740	1.8 (0.2)	507	1.7 (0.4)	154	1.9 (0.4)
<i>Nursing Home/Assisted Living</i>	17,045	40.5 (0.2)	3,944	13.5 (0.4)	306	3.8 (0.1)
<i>Correctional Facility</i>	2,996	7.1 (0.7)	3,516	12.0 (0.1)	240	3.0 (0.4)
<i>Group Home</i>	1,462	3.5 (0.5)	866	3.0 (0.4)	61	0.8 (0.1)
<i>Emergency Shelter</i>	478	1.1 (0.3)	107	0.36 (0.04)	6	0.1 (0.1)
<i>Some Other GQ</i>	2,537	6.0 (0.9)	1,818	6.21 (0.04)	210	2.6 (0.2)
<i>Move</i>	1,766	4.2 (0.4)	391	1.34 (0.01)	117	1.5 (0.1)
<i>Multiple Living Situations Mentioned</i>	4,187	9.96 (0.01)	1,745	6.0 (0.1)	359	4.5 (0.6)
<i>Don't Know/Roster Duplicate</i> <sup>37</sup>	626	1.5 (0.3)	241	0.8 (0.1)	34	0.4 (0.2)
<b>Total</b>	<b>108,473</b>	<b>100.0</b>	<b>59,201</b>	<b>100.0</b>	<b>20,149</b>	<b>100.0</b>

Source: Unduplication CFU Analysis Files

For within-county links that mentioned a complex living situation, respondents mentioned staying at a Nursing Home or Assisted Living facility 40.5 percent of the time. Respondents also mentioned a college living situation 20.9 percent of the time.

<sup>37</sup> After verifying the address of the CFU household, the interviewer next verified the roster from the original enumeration. During this roster verification, a respondent indicated that they did not know someone on the roster or that the person was already listed elsewhere on the roster.

For within-state links that mentioned a complex living situation, respondents mentioned being away for college 51.9 percent of the time. Respondents also mentioned staying at a Nursing Home/Assisted Living facility and staying at a correctional facility, 13.5 percent and 12.0 percent, respectively.

For across-state links that mentioned a complex living situation, respondents mentioned being away for college more than any other reason at 72.7 percent of the time.

Across all three geographies, respondents consistently mentioned being away for college and the percentage increasingly grew with the geographic distance between returns. The other two living situations mentioned frequently were staying at a Nursing Home or Assisted Living and staying at a correctional facility. A person may know they were staying at a Nursing Home or Assisted Living, but still have a home where they consider themselves living.

#### **5.2.4 Resolving Duplication<sup>38</sup>**

After a respondent indicated that a person in the CFU household had a complex living situation, the CFU instrument attempted to determine where the person lived or stayed most of the time. After the interview was completed, additional processing then determined whether the person remained as a resident at the CFU household or was deleted from the roster according to the census residence rule. For example, when a respondent indicated that the duplicated person was staying at a nursing home on Census Day, then the duplicated person was deleted from the household roster, since the Census Residence Rule states they should be counted at the nursing home. The residence rule and residence situations are included as Appendix A. It should be noted that the residence rule is only applied to the CFU return that was contacted and completed the interview.

As stated earlier, this analysis includes cases that were considered both production and evaluation. Production cases ultimately affected Census counts, but evaluation cases did not. There is no distinction made between the two in these analyses because the goal was to determine how effective the current CFU interview would perform with duplication cases in general.

---

<sup>38</sup> Note that this analysis includes cases that were considered both production and evaluation. Production cases ultimately affected 2010 Census counts, but evaluation cases did not. There is no distinction made between the two in these analyses because the goal was to determine how effective the current CFU interview would perform with duplication cases in general.

### 5.2.4.1 Deletion Rates

Ultimately, if the CFU instrument could resolve duplication perfectly, then a person record on one side of each person link should be deleted from a housing unit. Table 59 shows the number and percent of person links that contain a delete. This is shown for the entire universe and for links where a complex situation was mentioned.

**Table 59: Deletion Rate of Person Links**

<b>Resolution<sup>39</sup></b>	<b>Complex Situation Mentioned</b>		<b>Entire Universe</b>	
	<b>Number</b>	<b>Percent (Std. Error)</b>	<b>Number</b>	<b>Percent (Std. Error)</b>
Link does not contain a deleted person	606,950	46.8 (0.1)	1,519,639	68.8 (0.1)
Link contains a deleted person	689,920	53.2 (0.1)	689,923	31.2 (0.1)
<b>Total</b>	<b>1,296,870</b>	<b>100.0</b>	<b>2,209,562</b>	<b>100.0</b>

Source: Unduplication CFU Analysis Files

Table 59 indicates that when looking at the entire universe, only 31.2 percent of person links resulted in at least one side of a person link being deleted. Ideally, this percent would be close to 100, since at least one side should have a deleted person. However, when a complex situation was mentioned, the percent of person links with a deletion increased to 53.2 percent.

Deletions generally do not occur without first acknowledging a complex living situation. In Table 59, the number of person links that contain a deleted person from the entire universe is 689,923. However, this number falls only to 689,920 when the universe is restricted to links that had mentioned a complex living situation. There were three links that deleted a person without mentioning a complex situation. Two of those persons were not data defined, and one person was identified as a person born after Census Day. Both situations result in a person being deleted from a household without having mentioned a complex living situation.

Table 60 shows the number and percent of duplicate person links that contain a delete by stratum for the HU to HU person links. One of the goals of this evaluation was to determine whether unduplication-only cases would perform better in a true census environment and to see whether long distance duplication (within state and across state) would perform well, since the Census tests were mostly within county links.

<sup>39</sup> After the CFU production period ended, an unexpected logical pathway was discovered in the algorithm that determined if a roster member should be kept, added, or removed from a household. Due to incorrect programming, all persons who should have been deleted from a roster due to being in a correctional facility, an emergency shelter, a group home, or some other group quarter on Census Day remained residents. Only if a person was identified for deletion for one of these reasons in conjunction with another reason (e.g., college, nursing home, etc.) were they successfully deleted from the roster. This error affected the roster change rate for the overcount strata “in jail or prison”.

**Table 60: Deletion Rates by Stratum – Entire Universe of HU to HU Person Links**

Strata	Does Not Contain a Delete		Contains a Delete		Row Total	
	Number	Percent (Std. Error)	Number	Percent (Std. Error)	Number	Percent
OC - College HU-HU	43,858	39.9 (0.4)	66,035	60.1 (0.4)	109,894	100.0
OC - Nursing Home HU-HU	1,730	55.0 (0.2)	1,415	45.0 (0.2)	3,146	100.0
OC – Jail HU-HU <sup>40</sup>	2,533	90.4 (0.1)	270	9.6 (0.1)	2,804	100.0
OC - Military HU-HU	6,630	74.5 (1.1)	2,274	25.5 (1.1)	8,904	100.0
OC - Other Above HU-HU	137,408	67.8 (0.5)	65,397	32.2 (0.5)	202,805	100.0
OC - Seasonal HU-HU	315,808	56.0 (0.2)	247,684	44.0 (0.2)	563,492	100.0
OC - Custody HU-HU	236,542	72.7 (0.4)	88,802	27.3 (0.4)	325,343	100.0
OC - Other Below HU-HU	160,246	78.8 (0.6)	43,050	21.2 (0.6)	203,296	100.0
Within County HU-HU	261,298	93.0 (0.1)	19,747	7.0 (0.1)	281,046	100.0
Within State HU-HU	100,861	89.86 (0.02)	11,376	10.14 (0.02)	112,237	100.0
Across State HU-HU	78,551	93.1 (0.2)	5,830	6.9 (0.2)	84,381	100.0
<b>Total</b>	<b>1,345,466</b>	<b>70.9 (0.1)</b>	<b>551,881</b>	<b>29.1 (0.1)</b>	<b>1,897,347</b>	<b>100.0</b>

Source: Unduplication CFU Analysis Files

Table 60 shows that the overcount categories that have generally performed well in the CFU instrument (college, nursing home, military and other above) (Poehler, 2010b) are all generally performing better than the rest of the person links. The overcount jail category would have performed better if not for the coding error mentioned in the footnote to Table 59, so a number of persons who should have been deleted from the roster remained residents. For CFU, the overcount seasonal category was not considered a production case because it had performed poorly during the mid-decade tests. However, Table 60 shows that it was one of the most successful strata (44.0 percent of links contained a deletion). This indicates that persons with seasonal or second homes may respond better in an actual census environment when duplication can be identified nationally, or when the combination of duplication and the seasonal overcount box are taken into consideration jointly. The workload of cases from the seasonal stratum is significant, however (they are the largest portion of cases in Table 60), so the utility of those cases has significant cost impacts.

The most disappointing result from Table 60 is that the unduplication-only strata appear to be performing very poorly. Within county HU-HU person links were linked to a delete only 7.0 percent of the time, within state HU-HU person links were linked to a delete only 10.1 percent of the time, and across state HU-HU person links were linked to a delete only 6.9 percent of the time. This is not surprising since these strata had lower occurrences of complex living situations being mentioned, as shown in Section 5.2.3.1.

<sup>40</sup> This number of OC – Jail HU-HU deletes would be higher if the residence coding logic had been successfully implemented as intended.

Table 61 shows the number and percent of potential duplicates linked to a delete by stratum for HU to GQ links. Again, one of the goals of this evaluation was to determine whether unduplication-only cases would perform well since the 2006 Census Test was not able to adequately test the identification of HU to GQ duplication.

**Table 61: Deletion Rates by Stratum – Entire Universe of HU to GQ Links**

Strata	Does Not Contain a Delete		Contains a Delete		Row Total	
	Number	Percent (Std. Error)	Number	Percent (Std. Error)	Number	Percent
OC - College HU-GQ	8,882	10.9 (0.3)	72,485	89.1 (0.3)	81,367	100.0
OC - Nursing Home HU-GQ	1,506	23.6 (0.1)	4,875	76.4 (0.1)	6,381	100.0
OC – Jail HU-GQ <sup>41</sup>	3,980	93.8 (0.3)	265	6.2 (0.3)	4,245	100.0
OC - Military HU-GQ	651	38.4 (0.1)	1,044	61.6 (0.1)	1,695	100.0
OC - Other Above HU-GQ	4,174	30.6 (0.3)	9,466	69.4 (0.3)	13,640	100.0
OC - Seasonal HU-GQ	2,753	75.4 (3.2)	899	24.6 (3.2)	3,652	100.0
OC - Custody HU-GQ	808	79.5 (9.5)	208	20.5 (9.5)	1,016	100.0
OC - Other Below HU-GQ	9,570	77.2 (2.3)	2,825	22.8 (2.3)	12,395	100.0
Within County HU-GQ	84,864	78.2 (0.1)	23,609	21.8 (0.1)	108,473	100.0
Within State HU-GQ	42,317	71.5 (0.4)	16,884	28.5 (0.4)	59,200	100.0
Across State HU-GQ	14,668	72.8 (0.4)	5,483	27.2 (0.4)	20,151	100.0
<b>Total</b>	<b>174,173</b>	<b>55.8 (0.2)</b>	<b>138,042</b>	<b>44.2 (0.2)</b>	<b>312,215</b>	<b>100.0</b>

Source: Unduplication CFU Analysis Files

We see that the four of the CFU production overcount categories are performing very well. The CFU evaluation overcount cases are not performing very well, but the non-production overcount categories are not typically associated with GQs, the universe for this table.

The unduplication-only HU-GQ links also did not perform that well, showing a low of 21.8 percent of duplicates linked to a delete for within county links and a high of 28.5 percent of duplicates linked to a delete for within state links. Once again, however, they are performing better than their HU-HU counterparts.

The following tables will show the change in deletion rate when a complex situation is mentioned. Table 59 had presented the deletion rates for the entire universe and the deletion rates for the universe of cases that had mentioned a complex situation. Table 62 expands on the deletion rate for cases that mentioned a complex living situation, delineating between HU-HU or HU-GQ links.

<sup>41</sup> This number of OC – Jail HU-GQ deletes would be higher if the residence coding logic had been successfully implemented as intended.

**Table 62: Deletion Rate for Person Links that Mentioned a Complex Situation, by Type of Link**

	HU-HU Links		HU-GQ Links		Overall	
	Number	Percent (Std. Error)	Number	Percent (Std. Error)	Number	Percent (Std. Error)
<b>Resolution</b> <sup>42</sup>						
Link does not contain a deleted person	557,642	50.3 (0.2)	49,307	26.3 (0.2)	606,950	46.8 (0.1)
Link contains a deleted person	551,878	49.7 (0.2)	138,042	73.7 (0.2)	689,920	53.2 (0.1)
<b>Total</b>	<b>1,109,520</b>	<b>100.0</b>	<b>187,350</b>	<b>100.0</b>	<b>1,296,870</b>	<b>100.0</b>

Source: Unduplication CFU Analysis Files

HU-GQ links that mentioned a complex living situation deleted a person 73.7percent of the time, compared to HU-HU links that deleted a person 49.7percent of the time. The residence rule for GQs is based on April 1 location, so it is easier to apply than the residence rule for housing units that has to determine the location where an individual is most of the time.

Table 63 shows the number and percent of potential duplicates containing a delete by stratum for HU-HU links when a complex situation was mentioned.

<sup>42</sup>After the CFU production period ended, an unexpected logical pathway was discovered in the algorithm that determined if a roster member should be kept, added, or removed from a household. Due to incorrect programming, all persons who should have been deleted from a roster due to being in a correctional facility, an emergency shelter, a group home, or some other group quarter on Census Day remained residents. Only if a person was identified for deletion for one of these reasons in conjunction with another reason (e.g., college, nursing home, etc.) were they successfully deleted from the roster. This error affected the roster change rate for the overcount strata “in jail or prison”.



**Table 63: Deletion Rate for HU-HU Person Links that Mentioned a Complex Situation, by Stratum**

Strata	Does not Contain a Delete		Contains a Delete		Row Total	
	Number	Percent (Std. Error)	Number	Percent (Std. Error)	Number	Percent
OC - College HU-HU	28,061	29.82 (0.05)	66,035	70.2 (0.05)	94,096	100.0
OC - Nursing Home HU-HU	546	27.8 (0.6)	1,415	72.18 (0.6)	1,961	100.0
OC – Jail HU-HU <sup>43</sup>	519	65.7 (0.1)	270	34.3 (0.1)	789	100.0
OC - Military HU-HU	1,828	44.6 (1.7)	2,273	55.4 (1.7)	4,101	100.0
OC - Other Above HU-HU	69,782	51.6 (0.6)	65,396	48.4 (0.6)	135,178	100.0
OC - Seasonal HU-HU	180,900	42.2 (0.2)	247,683	57.8 (0.2)	428,583	100.0
OC - Custody HU-HU	167,417	65.3 (0.3)	88,802	34.7 (0.3)	256,219	100.0
OC - Other Below HU-HU	46,218	51.8 (1.2)	43,050	48.2 (1.2)	89,269	100.0
Within County HU-HU	37,148	65.3 (0.9)	19,747	34.7 (0.9)	56,895	100.0
Within State HU-HU	17,063	60.0 (0.2)	11,376	40.0 (0.2)	28,439	100.0
Across State HU-HU	8,162	58.3 (0.2)	5,830	41.7 (0.2)	13,992	100.0
<b>Total</b>	<b>557,642</b>	<b>50.3 (0.2)</b>	<b>551,878</b>	<b>49.7 (0.2)</b>	<b>1,109,520</b>	<b>100.0</b>

Source: Unduplication CFU Analysis Files

Table 60 showed that across-state HU-HU duplication resulted in 6.9 percent of links being linked to a duplicate, but Table 63 shows that the percentage increases to 41.7 percent when the respondent acknowledged a complex living situation. The custody stratum has one of the lowest deletion rates within cases that mentioned a complex living situation (34.7 percent of the time). Deleting a duplicated person from a housing unit requires the respondent to acknowledge that the duplicated person spends more time at the other address than at their own address, which parents involved in custody arrangements are not prone to do (qualitative evidence of this can be seen in Section 5.4.5).

Table 64 shows the number and percent of potential duplicates containing a delete by stratum for HU-GQ links when a complex situation was mentioned.

<sup>43</sup> This number of OC – Jail HU-HU deletes would be higher if the residence coding logic had been successfully implemented as intended.

**Table 64: Deletion Rate for HU-GQ Person Links that Mentioned a Complex Situation, by Stratum**

Strata	Does not Contain a Delete		Contains a Delete		Row Total	
	Number	Percent (Std. Error)	Number	Percent (Std. Error)	Number	Percent
OC - College HU-GQ	5,846	7.5 (0.2)	72,485	92.5 (0.2)	78,331	100.0
OC - Nursing Home HU-GQ	420	7.9 (0.4)	4,875	92.1 (0.4)	5,295	100.0
OC – Jail HU-GQ <sup>44</sup>	3,018	91.9 (0.4)	265	8.1 (0.4)	3,282	100.0
OC - Military HU-GQ	196	15.8 (0.2)	1,044	84.2 (0.2)	1,241	100.0
OC - Other Above HU-GQ	2,329	19.75 (0.1)	9,466	80.3 (0.1)	11,796	100.0
OC - Seasonal HU-GQ	950	51.4 (1.4)	899	48.6 (1.4)	1,849	100.0
OC - Custody HU-GQ	365	63.7 (8.1)	208	36.3 (8.1)	573	100.0
OC - Other Below HU-GQ	2,817	49.9 (4.1)	2,825	50.1 (4.1)	5,642	100.0
Within County HU-GQ	18,428	43.84 (0.02)	23,609	56.2 (0.02)	42,037	100.0
Within State HU-GQ	12,377	42.3 (0.6)	16,884	57.7 (0.6)	29,261	100.0
Across State HU-GQ	2,561	31.8 (0.7)	5,483	68.2 (0.7)	8,044	100.0
<b>Total</b>	<b>49,307</b>	<b>26.3 (0.2)</b>	<b>138,042</b>	<b>73.7 (0.2)</b>	<b>187,350</b>	<b>100.0</b>

Source: Unduplication CFU Analysis Files

The HU-GQ unduplication-only strata are more successful than the HU-HU unduplication-only strata. Over 50 percent of all HU-GQ unduplication-only strata deleted a person while none of the HU-HU unduplication-only strata from Table 63 had a success rate over 50 percent.

#### 5.2.4.2 Residency Status

Another way to investigate the resolution of the duplication is to examine the final residency status of both sides of a link. Each person who goes through the CFU interview will be given a status of resident or non-resident (delete). If a CFU interview was not completed, then a final residency cannot be determined and that person's status is unresolved.

Before a CFU interview actually begins, one of the first modules in the CFU instrument verifies that we have reached the correct address. If the respondent indicates that we have not reached the correct address, and that they were not living at the address in question on Census Day, then the CFU interview would end without collecting further information. This could happen if a household or persons moved around Census Day. We refer to those cases as Mod C cases (named after the CFU module that verifies the address).

Table 65 shows the linked residence status when controlling for whether a HU to HU link completed one or two interviews.

<sup>44</sup> This number of OC – Jail HU-GQ deletes would be higher if the residence coding logic had been successfully implemented as intended.

**Table 65: CFU Residence Status for HU-HU Person Links**

<b>Linked Residency Status</b>	<b>Number</b>	<b>Percent (Std. Error)</b>
<b>One side completed a CFU interview</b>	<b>997,459</b>	<b>100.0</b>
Resident - Unresolved	804,624	80.7 (0.1)
Delete - Unresolved	168,616	16.9 (0.1)
Mod C - Unresolved	24,219	2.4 (0.1)
<b>Both sides completed a CFU interview</b>	<b>899,888</b>	<b>100.0</b>
Resident - Resident	462,140	51.4 (0.3)
Resident - Delete	364,812	40.5 (0.2)
Resident - Mod C	53,513	5.9 (0.1)
Resident - Unresolved	68	0.008 (0.001)
Delete - Delete	13,779	1.53 (0.05)
Delete - Mod C	4,657	0.52 (0.02)
Delete - Unresolved	16	0.0018 (0.0003)
Mod C - Mod C	898	0.10 (0.01)
Mod C - Unresolved	4	0.0004 (0.0002)
<b>Total</b>	<b>1,897,347</b>	<b>100.0</b>

Source: Unduplication CFU Analysis Files

Table 65 shows that if we completed interviews with both sides of the person link, we deleted one side of the link 40.5 percent of the time, compared to 16.9 percent when we only completed an interview with one side of a link. This provides an indication that interviewing both sides is more successful at resolving duplication. However when we talked to both sides, there were 1.53 percent of person links in which both sides of the link were deleted. This is not an ideal situation, unless the person is involved with multiple links, because this would mean that the person would likely not be counted in the Census. This is possible though since the CFU interview with the one side of a link did not take into consideration any information obtained from the other side of the link. Additionally, in 51.4 percent of cases where we interviewed both sides, the individual remained a confirmed resident on both sides and was not deleted (remained duplicated). So even though completing an interview at both housing units resolves more duplication, it still leaves a less than ideal percentage of situations unresolved.

Table 66 shows the number and percent of each pair of resident statuses for HU to GQ links. Only the resident status of the CFU interview for the housing unit is shown since the CFU interview was not conducted with GQs.

**Table 66: CFU Residence Status for HU side of HU-GQ Person Links**

<b>Residence Status for Housing Unit Side</b>	<b>Number</b>	<b>Percent</b>
Resident	169,150	54.2 (0.1)
Delete	138,042	44.2 (0.2)
Mod C	5,022	1.6 (0.1)
<b>Total</b>	<b>312,215</b>	<b>100.0</b>

Source: Unduplication CFU Analysis Files

Table 66 indicates that most persons in HU-GQ links are residents (54.2 percent) but 1.6 percent of all HU-GQ person links indicated they were not living at the CFU address on Census Day (Mod C Cases). It should be investigated whether this information can be used as an indication that the person can be removed from the HU and only counted at the GQ. The census residence rule states that people in certain types of GQs on Census Day should be counted at the GQ.

#### 5.2.4.3 Deletion Rates of Characteristics and Demographic Characteristics of Person Links

The following tables discuss the deletion rates of certain characteristics of the CFU universe and contain the 2,209,562 person links in CFU.

##### Phone match

Table 67 shows how often person links in CFU contained a deleted person, by phone match status.

**Table 67: Deletion Rate of Person Links by Phone Match Status**

Phone Match	Person Links Containing A Delete		Total CFU Universe
	Number	Percent (Std. Error)	Number
Yes	109,274	42.6 (0.3)	256,756
No	397,776	28.3 (0.1)	1,407,803
One Blank	42,163	19.0 (0.3)	222,258
Both Blank	2,668	25.3 (1.6)	10,530
GQ	138,042	44.2 (0.2)	312,215
<b>Total</b>	<b>689,923</b>	<b>31.2 (0.1)</b>	<b>2,209,562</b>

Source: Unduplication CFU Analysis Files

HU-GQ links were the most likely to contain a deleted person (44.2 percent), followed closely by phone match links, which had a deleted person 42.6 percent of the time.

##### Geography

Table 68 shows how often person links in CFU contained a deleted person, by geography. This includes overcount and unduplication-only strata.

**Table 68: Deletion Rate of Person Links by Geography**

<b>Geography</b>	<b>Person Links Containing A Delete</b>		<b>Total CFU Universe</b>
	<b>Number</b>	<b>Percent (Std. Error)</b>	<b>Number</b>
Within County	201,467	19.7 (0.2)	1,022,890
Within State	298,010	41.0 (0.2)	726,640
Across State	190,446	41.4 (0.2)	460,032
<b>Total</b>	<b>689,923</b>	<b>31.2 (0.1)</b>	<b>2,209,562</b>

Source: Unduplication CFU Analysis Files

Within-county person links were the least likely to contain a deleted person (19.7 percent), while both within-state and across-state links had over forty percent of links with a deletion.<sup>45</sup>

### Age

Table 69 shows how often person links in CFU contained a deleted person, by age.

<sup>45</sup> Additional analysis affirmed that more deletions occurred in the overcount strata than in the unduplication-only strata.

**Table 69: Deletion Rate of Person Links by Age**

<b>Age</b>	<b>Person Links Containing A Delete</b>		<b>Total CFU Universe</b>
	<b>Number</b>	<b>Percent (Std. Error)</b>	<b>Number</b>
Under 5 years	17,176	15.4 (0.6)	111,847
5 to 9 years	43,509	24.4 (0.6)	178,592
10 to 14 years	58,642	25.2 (0.4)	232,616
15 to 19 years	106,851	36.5 (0.3)	293,003
20 to 24 years	134,017	44.2 (0.2)	303,080
25 to 29 years	16,853	14.2 (0.2)	118,702
30 to 34 years	9,896	13.6 (0.3)	72,928
35 to 39 years	10,892	16.7 (0.7)	65,392
40 to 44 years	14,953	20.4 (0.8)	73,417
45 to 49 years	24,624	26.0 (0.4)	94,723
50 to 54 years	34,631	31.4 (0.6)	110,266
55 to 59 years	46,781	38.0 (0.9)	122,964
60 to 64 years	51,421	42.0 (0.3)	122,404
65 to 69 years	41,834	43.3 (0.3)	96,666
70 to 74 years	27,307	40.1 (0.3)	68,091
75 to 79 years	18,707	39.4 (1.3)	47,447
80+ years	27,996	38.1 (0.6)	73,413
Contradictory	3,830	16.0 (1.2)	23,981
Missing	2	6.5 (1.4)	31
<b>Total</b>	<b>689,923</b>	<b>31.2 (0.1)</b>	<b>2,209,562</b>

Source: Unduplication CFU Analysis Files

The under 5 years and 25 to 39 year old person links were the least likely to contain a deleted person, while 20 to 24 years old (generally college-aged) and 55 years and over person links were the most likely to contain a delete.

## Sex

Table 70 shows how often person links in CFU contained a deleted person, by sex.

**Table 70: Deletion Rate of Person Links by Sex**

	Person Links Containing A Delete		Total CFU Universe
Sex	Number	Percent (Std. Error)	Number
Male	343,741	29.9 (0.2)	1,149,086
Female	345,026	32.7 (0.1)	1,055,383
Contradicting	1,151	23.1 (2.3)	4,986
Missing	5	5.1 (2.4)	107
<b>Total</b>	<b>689,923</b>	<b>31.2 (0.1)</b>	<b>2,209,562</b>

Source: Unduplication CFU Analysis Files

Female person links were the most likely to contain a deleted person (32.7 percent of the time), while male links had 29.9 percent of links with a deletion.

### Hispanic Origin

Table 71 shows how often person links in CFU contained a deleted person, by Hispanic origin.

**Table 71: Deletion Rate of Person Links by Hispanic Origin**

	Person Links Containing A Delete		Total CFU Universe
Hispanic Origin	Number	Percent (Std. Error)	Number
Not Hispanic or Latino checkbox only	640,196	32.8 (0.1)	1,949,118
Mexican checkbox only	17,605	17.8 (0.2)	98,800
Puerto Rican checkbox only	3,634	22.7 (1.7)	16,025
Cuban checkbox only	1,193	13.6 (1.8)	8,801
Another Hispanic checkbox only	88	12.5 (2.4)	703
Multiple checkboxes	255	35.8 (7.0)	712
Both Checkbox and Write-in	6,619	22.4 (0.6)	29,610
Write-in Only	298	20.1 (4.3)	1,479
Contradictory	19,717	19.1 (0.1)	103,377
Missing	317	33.9 (3.7)	935
<b>Total</b>	<b>689,923</b>	<b>31.2 (0.1)</b>	<b>2,209,562</b>

Source: Unduplication CFU Analysis Files

Non-Hispanic links (and the few cases with multiple checkboxes and missing Hispanic origin) were more likely to contain a deleted person (32.8 percent, 35.8 percent and 33.9 percent, respectively). Cuban and another Hispanic checkbox only person links were the least likely to contain a deleted person (13.6 percent and 12.5 percent, respectively).

## Race

Table 72 shows how often person links in CFU contained a deleted person, by race.

**Table 72: Deletion Rate of Person Links by Race**

Race	Person Links Containing A Delete		Total CFU Universe
	Number	Percent (Std. Error)	Number
White checkbox alone	559,299	34.7 (0.1)	1,612,481
Black or African American checkbox alone	46,098	20.1 (0.6)	229,389
American Indian and Alaska Native checkbox alone	76	14.7 (6.1)	515
Asian Indian checkbox alone	4,369	35.5 (2.0)	12,319
Chinese checkbox alone	8,685	34.5 (2.8)	25,157
Filipino checkbox alone	3,739	26.1 (0.9)	14,313
Japanese checkbox alone	1,534	40.6 (2.7)	3,783
Korean checkbox alone	2,244	29.7 (2.2)	7,566
Vietnamese checkbox alone	1,817	24.6 (1.4)	7,392
Native Hawaiian checkbox alone	99	28.7 (7.5)	345
Guamanian or Chamorro checkbox alone	68	50.8 (15.9)	135
Samoan checkbox alone	32	14.5 (7.6)	218
Other Asian checkbox alone	30	50.5 (12.6)	59
Other Pacific Islander checkbox alone	0	0.0 (0.0)	3
Some Other Race checkbox alone	39	14.7 (7.6)	267
Multiple checkboxes	4,764	28.5 (1.7)	16,691
Both Checkbox and Write-in	15,918	24.3 (1.1)	65,528
Write-in Only	799	16.7 (2.7)	4,771
Contradictory	40,094	19.3 (0.3)	207,850
Missing	220	28.3 (4.0)	779
<b>Total</b>	<b>689,923</b>	<b>31.2 (0.1)</b>	<b>2,209,562</b>

Source: Unduplication CFU Analysis Files

Table 72 shows a wide range of deletion rates across race.

## 5.3 Experimental CFU Questions (Mod Q Cases)

There are cases where the CFU respondent does not allude to or acknowledge a complex living situation as was indicated on the original census return. For instance, the initial census return could show that a child lived or stayed at another place for a custody situation, but the CFU respondent might never mention a custody situation.

One potential reason for this is that the CFU interview might not provide sufficient cues for respondents to add or delete people from their initial household roster. An experiment was



designed as part of the 2010 Census to test additional probing questions that could enhance future followup operations. During the 2010 CFU operation, a series of experimental questions (called Mod Q) was administered as an exit interview to a sample of respondents who did not mention the complex living situation that we expected them to mention during the standard CFU interview. The text in Mod Q specifically mentioned what was marked on the original census return (which the current CFU interview does not do) and asked if the respondent remembered what he or she was thinking about when he or she answered the original census questionnaire. The experimental questions were cognitively tested before being implemented. For more information please see Childs, et al (2009).

A sample of cases was identified as eligible for Mod Q before being sent to CFU. During the CFU interview, if there was no mention of the expected living situation expected for those sampled cases based on the overcount or undercount boxes originally marked on that case's census form, then the CFU respondent was asked the Mod Q questions. Mod Q probed specifically on the overcount or undercount categories that had been marked on the original census questionnaire. Responses to the experimental questions were not used as an official 2010 Census response.

Mod Q had two sections: Mod Q undercount questions and Mod Q overcount questions. The Mod Q undercount questions probed for missing persons, specifically referencing the undercount category that had been marked on the original census form. Mod Q undercount cases are not a focus for this evaluation.

The Mod Q overcount questions probed for the living situations of overcounted persons, specifically referencing the overcount category that had been marked on the original census form. If any names on the roster were confirmed to have another place to stay, they were referred to as Mod Q deletes. If the Mod Q deletes stayed somewhere other than the census address, then the interviewer probed for that alternative address, where the Mod Q delete spent most of their time in March and April 2010, and how much time the Mod Q delete spent at each address in the last 12 months. The remainder of section 5.3 will focus on these Mod Q overcount persons.

### **5.3.1 Universe for Mod Q**

Section 5.2 analyzed cases as links; however, this section will present results only for individual persons identified as duplicates who were eligible for Mod Q. While it might be ideal to analyze them at the link level, one of the goals of Mod Q was to determine why a person marked an overcount box at the individual level but did not mention the same living situation in CFU.

Table 73 shows the number and percent of persons eligible for Mod Q and how the CFU interview was completed.

**Table 73: Eligible and Completed Mod Q Persons**

<b>CFU/Mod Q status</b>	<b>Number</b>	<b>Percent</b>
Interviews Resolved In CFU	59,295	47.3
Sufficient Partial or Mod C	1,603	1.3
Interviews Sent to Mod Q and No Data	384	0.3
Undercount Only	3,423	2.7
Overcount Household Multiple	31,117	24.8
Interviews Sent to Mod Q and Completed	29,548	23.6
<b>Total CFU Completes Eligible for Mod Q</b>	<b>125,370</b>	<b>100.0</b>

Source: CFU Analysis Files

Of the 125,370 potentially duplicated persons eligible for Mod Q, 47.3 percent addressed the overcount coverage issue in CFU, making Mod Q irrelevant. There were 1,603 persons that exited CFU prematurely and were classified as a sufficient partial or they exited the interview through Mod C, so they were not sent to Mod Q. An additional 384 persons were identified as being sent to Mod Q but no data came back for them. These appear to be cases that went into Mod Q and exited the interview. There were 3,423 unduplication persons that completed a CFU interview and were sampled for Mod Q for an undercount reason and are out of scope for this study. There were 31,117 unduplication persons that completed a CFU interview and were sampled for Mod Q because multiple persons in the household had marked various overcount reasons, and these accounted for nearly one-quarter of all cases eligible for Mod Q. The remaining 29,548 persons were sent to Mod Q and completed that module.

The 3,423 undercount-only cases are not expected to provide any information to explain the duplication so they are omitted from this analysis. Additionally the Overcount Household Multiple cases did not go to Mod Q because the system could not handle the multiple categories, so they are also omitted from this analysis. Thus, excluding the 3,423 undercount-only persons that did not mark an overcount category and the 31,117 household multiple persons, 90,830 persons were eligible for the overcount series in Mod Q. Table 74 shows how many duplicate persons completed Mod Q by the overcount category.

**Table 74: Fate of Duplicate Persons Eligible for Mod Q, by Initial Overcount Categories**

<b>Overcount Reason</b>	<b>Interviews Resolved In CFU</b>		<b>Sufficient Partial or Mod C</b>		<b>Interviews Sent to Mod Q and No Data</b>		<b>Interviews Sent to Mod Q and Completed</b>		<b>Total</b>	
	<b>Number</b>	<b>Percent</b>	<b>Number</b>	<b>Percent</b>	<b>Number</b>	<b>Percent</b>	<b>Number</b>	<b>Percent</b>	<b>Number</b>	<b>Percent</b>
College	14,039	93.4	327	2.2	10	0.1	651	4.3	15,027	100.0
Military	94	35.6	5	1.9	2	0.8	163	61.7	264	100.0
Jail	166	13.3	32	2.6	21	1.7	1,033	82.5	1,252	100.0
Nursing Home	303	21.5	73	5.2	28	2.0	1,006	71.3	1,410	100.0
Child Custody	24,988	80.5	332	1.1	94	0.3	5,641	18.2	31,055	100.0
Seasonal	18,229	66.6	513	1.9	111	0.4	8,507	31.1	27,360	100.0
Another Reason	1,476	10.2	321	2.2	118	0.8	12,547	86.8	14,462	100.0
<b>Total</b>	<b>59,295</b>	<b>65.3</b>	<b>1,603</b>	<b>1.8</b>	<b>384</b>	<b>0.4</b>	<b>29,548</b>	<b>32.5</b>	<b>90,830</b>	<b>100.0</b>

Source: CFU Analysis Files

There were 15,027 persons that marked the college overcount category and were identified as eligible for Mod Q that completed a CFU interview. Only 651, or 4.3 percent, completed Mod Q, which is an indication that respondents are mentioning this situation during the CFU interview quite often, and as supported by previously reported data. However, other categories were not resolved with such frequency in CFU so larger proportions of other categories flowed into Mod Q. For instance, 14,462 persons had marked the overcount category ‘for another reason’, completed a CFU interview, and been sampled for Mod Q. The majority of them (86.8 percent) completed Mod Q.

**5.3.2 Mod Q Overcount**

The Mod Q series of questions for overcount (OC) cases began with the following question, which was edited where italicized to apply to the specific overcount category that was applicable:

The Census Bureau is doing research about questions on the census form you completed earlier this year. At that time, you indicated that (NAME) sometimes lives or stays somewhere else *while in college housing*. Can you tell me what you were thinking about when you reported that?

- ( ) *away for college* in March or April 2010.
- ( ) *away for college* sometime in 2010, but not in March or April.
- ( ) *away for college* in 2009 or earlier
- ( ) away briefly *for college*
- ( ) *Enrolled in college or taking college course* but stays here
- ( ) Stays at another address, but not *for college*
- ( ) Another reason (please describe the reason below)

The response categories were not to be read aloud to the respondent. The interviewer was to mark one box and if the last box was marked, the interviewer was to enter a description in the textbox provided. The following table shows how often each response was recorded. (The table reflects all respondents and overcount situations, not just college situations from the text quoted above.)

Table 75: Reason Given In Mod Q for Marking the Initial Overcount Question		
Living Situation	Mod Q Overcount Persons	Percent
Stayed away for the stated OC category	7,702	26.1
<i>Away in March or April 2010</i>	2,760	35.8
<i>Away sometime in 2010</i>	1,263	16.4
<i>Away in 2009 or earlier</i>	719	9.3
<i>Away briefly</i>	2,960	38.4
(Engaged in OC category) but stays here	1,774	6.0
Stays at other address (not for OC category)	2,711	9.2
Another Reason	16,835	57.0
Don't Know	444	1.5
Refused	82	0.3
Total Mod Q Overcount Persons	29,548	100.0

Source: CFU Analysis Files

The most common reply to this question was ‘another reason’ (57.0 percent of the time). This could have happened either because the reply was unable to be coded by the interviewer into any of the given categories, or because the interviewers thought it easier to just type out the response. About 26.1 percent of respondents acknowledged that the

person stayed away from the CFU address for the Overcount category. Of that 26.1 percent, about 38.4 percent indicated having stayed away briefly, followed closely by 35.8 percent staying away in March and April 2010. Finally, 9.2 percent indicated staying at another address but not for the stated Overcount reason.

Table 76 below shows the living situation mentioned by overcount reason. This table was created to investigate if different overcount reasons responded differently.

Table 76: Reason Given In Mod Q for Marking the Initial Overcount Question, by Initial Overcount Mark

Living Situation	Initial Overcount Mark													
	OC – College		OC – Military		OC – Jail		OC – Nursing		OC – Child Custody		OC – Seasonal		OC – Another Reason	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent
Stayed away for the stated OC category	295	45.3	55	33.7	609	59.0	494	49.1	2,248	39.9	1,746	20.5	2,255	18.0
<i>Away in March or April 2010</i>	149	50.5	12	21.8	452	74.2	326	66.0	874	38.9	577	33.0	370	16.4
<i>Away sometime in 2010</i>	48	16.3	22	40.0	37	6.1	48	9.7	248	11.0	396	22.7	464	20.6
<i>Away in 2009 or earlier</i>	74	25.1	11	20.0	65	10.7	54	10.9	99	4.4	177	10.1	239	10.6
<i>Away briefly</i>	24	8.1	10	18.2	55	9.0	66	13.4	1,027	45.7	596	34.1	1,182	52.4
(Engaged in OC category) but stays here	62	9.5	18	11.0	54	5.2	61	6.1	233	4.1	605	7.1	741	5.9
Stays at other address (not for OC category)	20	3.1	5	3.1	15	1.5	28	2.8	96	1.7	906	10.7	1,641	13.1
Another Reason	258	39.6	84	51.5	330	31.9	390	38.8	2,984	52.9	5,063	59.5	7,726	61.6
Don't Know	12	1.8	1	0.6	19	1.8	28	2.8	66	1.2	154	1.8	164	1.3
Refused	4	0.6	0	0.0	6	0.6	5	0.5	14	0.2	33	0.4	20	0.2
<b>Total Mod Q Overcount Persons</b>	<b>651</b>	<b>100.0</b>	<b>163</b>	<b>100.0</b>	<b>1,033</b>	<b>100.0</b>	<b>1,006</b>	<b>100.0</b>	<b>5,641</b>	<b>100.0</b>	<b>8,507</b>	<b>100.0</b>	<b>12,547</b>	<b>100.0</b>

Source: CFU Analysis Files

Table 76 shows that college, jail and nursing home situations were most often able to indicate with a specific reason why they marked the question, as opposed to indicating ‘Another Reason’. All other categories relied on the ‘Another Reason’ category more than half of the time to explain why they marked the question. When a respondent answered ‘Another Reason’ we also captured an open-ended response for what they were thinking. The open-ended entries were not able to be analyzed in time for inclusion in this evaluation. Since the open-ended entries are such a large portion of the responses, it is difficult to interpret the results without them. It is recommended that the open-ended responses be revisited after they are coded.

If the respondent indicated that the person in question,

- stayed away in March/April 2010,
- stayed away sometime in 2010,
- stayed away briefly, or
- stayed at another address for a different reason,

then the interviewer was to collect the address of the place alluded to. The usefulness of the address collected is not assessed here, but it is recommended that they be further investigated. After asking for an address, the interviewer asked the respondent the following, known as the ‘most of the time’ or MOTT question:

**In March and April of this year, where did NAME spend most of the time?**

- ☐ **This address**
- ☐ **The other place**
- ☐ **Both places equally**

The interviewer was to read the response options to this question and mark only one box. This question was also asked in CFU and was used to determine which residence an individual should be counted at (the residence rule states that a person is to be counted where they spend most of their time). Since the results of Mod Q did not impact the final 2010 Census count, there was no question in this experimental series about a person’s location on April 1.

**Error! Not a valid bookmark self-reference.** shows how often each most-of-the-time response was provided, by the answer to the top-level question.

Table 77: Responses to the MOTT Question in Mod Q for Persons Away from the CFU Address

Location Most of the Time	Away in March or April 2010		Away Briefly		Away Sometime in 2010		Stays at Another Address		Total	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent
Census Address	808	29.3	2,064	69.7	1,013	80.2	1,172	43.2	5,057	52.2
Other Place	1,159	42.0	256	8.6	98	7.8	801	29.5	2,314	23.9
Both Places	751	27.2	592	20.0	138	10.9	684	25.2	2,165	22.3
Don't Know	14	0.5	15	0.5	5	0.4	27	1.0	61	0.6
Refused	2	0.1	4	0.1	0	0.0	1	0.0	7	0.1
Missing	26	0.9	29	1.0	9	0.7	26	1.0	90	0.9
Total	2,760	100.0	2,960	100.0	1,263	100.0	2,711	100.0	9,694	100.0

Source: CFU Analysis Files

There were 9,694 persons that went to this question in Mod Q. Just over half of the persons (52.2 percent) were reported to stay most of the time at the address we were calling. A little under one-quarter of the persons were reported to stay most of the time at the other address (23.9 percent), implying they should have been deleted from the roster of the address that was being interviewed. Persons described as being away in March and April 2010 were the least likely (29.3 percent) to be described as being at the census address most of the time in March and April of all top level responses.

Table 78 shows the most-of-the-time responses by the overcount category indicated.

Table 78: Responses to the MOTT Question in Mod Q for Persons Away from the CFU Address, by Overcount Mark

Location Most of the Time	OC – College		OC – Military		OC – Jail		OC – Nursing		OC – Child Custody		OC – Seasonal		OC-Another Reason	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent
Census Address	58	24.1	36	73.5	105	18.8	102	21.8	1,457	64.9	1,341	54.2	1,958	53.5
Other Place	126	52.3	6	12.2	387	69.2	300	64.1	83	3.7	545	22.0	867	23.7
Both Equally	52	21.6	7	14.3	53	9.5	52	11.1	687	30.6	549	22.2	765	20.9
Don’t Know	2	0.8	0	0.0	8	1.4	4	0.9	5	0.2	9	0.4	33	0.9
Refused	1	0.4	0	0.0	0	0.0	0	0.0	0	0.0	3	0.1	3	0.1
Missing	2	0.8	0	0.0	6	1.1	10	2.1	13	0.6	28	1.1	31	0.8
Total	241	100.0	49	100.0	559	100.0	468	100.0	2,245	100.0	2,475	100.0	3,657	100.0

Source: CFU Analysis Files

When looking at the most of the time question by overcount reason, we see that persons who marked the college, jail or nursing home category indicated that they were at the other address more than half of the time. All other categories indicated that the person was staying at the Census Address more than half of the time. Thus persons appear to be responding differently for HU and GQ type situations.

A last question was asked in the Mod Q interview:

**Please tell me how much time NAME spent at each of the addresses in the last 12 months.**  
(open text)

However, since this question was open-ended, the responses have not yet been compiled and analyzed. Again, once the open-ended text has been coded, it is recommended that these results be analyzed again.



## 5.4 Qualitative and Cognitive Research

A study was commissioned of RTI International and Research Support Services (RSS) by the Census Bureau using 2010 Census data to investigate the characteristics and questions of a followup interview that would prompt respondents to provide the information necessary to confirm and resolve person duplication, without anyone feeling that their privacy or confidentiality had been violated. Two substudies, one semi-structured and ethnographic in nature, called here “qualitative,” and one in the style of cognitive interviewing, were launched simultaneously to address the outlined objectives.

The interviews were conducted face-to-face with either suspected duplicates or a proxy respondent. Interviews were conducted in three primary sites: the Chicago metropolitan area, the Washington, DC, metropolitan area, and the Raleigh, NC metropolitan area; and two supplemental sites: the San Diego, CA metropolitan area and the New York City metropolitan area (only interviewed in the second round).

Cases were excluded for this study for various reasons that had made them in scope for another 2010 Census study:

- either side of the link marked the overcount question affirmatively on the initial census questionnaire,
- either side of the link was sent to CFU,
- either side of the link was in a Census Coverage Measurement (CCM) block, and
- either side of the link had been flagged as eligible for reinterview as part of the Census Program for Evaluations and Experiments (CPEX).

Information in this section was retrieved from the report “Qualitative Interviewing with Suspected Duplicates and Cognitive Testing of the Targeted Coverage Follow-up (TCFU) Interview” (Peytcheva, Sha, Cook, King, Gerber, Schoua-Glusberg, Kenward; 2011).

### 5.4.1 Objective

The specific aim of the qualitative study was to examine in-depth the household situations that cause duplication, from the point of view of the respondents themselves. A total of 50 interviews were carried out, in two rounds of 25 interviews each.

The goal of the cognitive study was to evaluate if the newly-developed TCFU questions functioned as intended (to confirm suspected duplications) and to detect if any concerns were raised by the respondent related to privacy and confidentiality in the interview. A total of 226 interviews were completed across two rounds of testing. The interviews covered 27 different classifications (nine of which involved GQ living arrangements).

## 5.4.2 Methodology

All interviews were conducted in a face-to-face mode with respondents who were either suspected duplicates from the 2010 Census or who were reporting for household members who were suspected duplicates. Respondents were paid \$40 after completing an interview.

A total of 46 cognitive and 25 qualitative interviews were conducted in fall 2010 (Round 1). This first round of interviews was small in order to identify any adjustments that needed to be made to the protocols or methodology. Some changes were made, mostly to recruiting protocols, after assessing those results and the rest of the interviews were conducted in spring and summer 2011 (Round 2). A total of 180 cognitive and 25 qualitative interviews were conducted in Round 2.

There were a number of important classifications and definitions used in this study. First, cases were classified into three distinct categories:

- If a person (or persons) was counted on two distinct housing unit returns, and the phone number was the same on each return, it was considered a Type 1 Case.
- If a person (or persons) was counted on two distinct housing unit returns, and the phone numbers were different, it was considered a Type 2 Case.
- If a person was counted in a GQ facility as well as in a housing unit, it was considered a Type 3 Case.

In Type 2 cases, where the two phone numbers were available, the researchers were also to attempt to interview both sides of a link and compare the answers. These interviews with both sides of a link were referred to as dependent interviews and will be discussed in Section 5.4.5.

The team of recruiters contacted over 3,000 households during the active recruitment period, with a success rate of 9.2 percent.

### 5.4.2.1 Cognitive Study

A scripted interview was used in the cognitive study (the TCFU instrument), though the interview contained different questions depending on whether the case was a Type 1, Type 2, or Type 3 case. The interview also contained questions for each suspected duplicate, making the interview longer for households with multiple duplicate persons. A structured cognitive protocol was administered after the TCFU questionnaire that guided the respondent through a retrospective think-aloud discussion of how they had answered the TCFU questions. The goal of the structured cognitive protocol was to assess the respondent's interpretation of the TCFU questions, gauge perception about the threat or sensitivity of any questions, and assess the overall burden in participating in the TCFU interview. In addition, respondents were debriefed on their experience with the 2010 Census and asked if they would be willing to provide the kind of information the TCFU was collecting over the phone. When time allowed, participants were asked to complete an Event History Calendar (EHC) that was designed to further cue recall and possibly close gaps in dates and activities reported in the TCFU. There were minimal differences between the protocol guides used in Round 1 and Round 2; most differences focused on

rewording and reorganizing some of the cognitive probes that were not understood well by respondents.

Suspected duplicates were classified based on their age, and questions for Type 1 and Type 2 cases were administered according to their age classification.

- Those aged 0-17 were classified as children.
- Those aged 18-70 were classified as adults.
- Those older than 70 were classified as seniors.

This classification allowed the TCFU interview questions to be tailored towards activities and living situations that were potentially relevant for that stage of life, as shown in Table 79.

**Table 79: Targeted Questions for TCFU, By Age Classification**

<b>Children</b>	<b>Adults</b>	<b>Seniors</b>
	Moved?	Moved?
	Attended college or grad school?	
	Been in the military?	Been in the military?
	Have a job?	Have a job?
	Seasonal or vacation home?	Seasonal or vacation home?
	Another residence owned or rented?	Another residence owned or rented?
Stay with a parent?	Stay with a parent?	
Stay with a grandparent?		
	Stay with son or daughter?	Stay with son or daughter?
	Stay with brother or sister?	Stay with brother or sister?
Stay with another relative?	Stay with another relative ?	Stay with another relative?
	Stay with boyfriend, girlfriend, or significant other?	Stay with companion, or significant other?
Stay with a friend, neighbor, or legal guardian?	Stay with friends?	Stay with friends?
Have a different address for school?		
Stay elsewhere for childcare purposes?		
		Nursing home, assisted living, independent living stay?
		Rehabilitation hospital stay?
		Stay elsewhere for health reasons?
Foster child?		
Other reasons?	Other reasons?	Other reasons?

Source: TCFU Training Materials for Contractors

Households in Type 1 or Type 2 cases were also evaluated to determine if they were a whole-household match, meaning that every single person listed on one linked questionnaire also was listed on the other linked questionnaire, and vice versa (in line with the analysis done in Section 5.1.6). All other links were considered to be partial household matches.

TCFU interviews were to be distributed across the combination of Type 1, Type 2, and Type 3 cases. Within Type 1 and Type 2 cases, interviews were to be distributed across both whole household matches and partial household matches, and with a variety of ages reflected in the duplicated persons. Within Type 3 cases, interviews were to be distributed across different types of GQs in which persons could be duplicated.

In an attempt to improve the CFU instrument, the Census Bureau adjusted the methodology used in the CFU interview and also developed a new set of follow-up questions while creating the TCFU questionnaire. The differences between the 2010 CFU interview and the tested TCFU interview are shown in Table 80.

.

**Table 80: Differences between the CFU and TCFU interviews**

<b>CFU</b>	<b>TCFU</b>
Attempted to speak with the person who filled out the census questionnaire	Attempted to speak with the duplicated person, if an adult
Mentioned the reported roster of household members as collected by the 2010 Census	Did not review the roster
Asked if anyone should be added to the roster	Did not attempt to add anyone to the roster
Asked whether “anyone” stayed at certain other locations <ul style="list-style-type: none"> <li>• Moved</li> <li>• Attended College</li> <li>• Child staying with someone else</li> <li>• Away in the military</li> <li>• For a job or business</li> <li>• Have a seasonal or second home</li> <li>• Stays elsewhere for any other reason</li> </ul>	Asked specifically whether the duplicated person(s) stays at certain locations. Options changed based on the age of the duplicated person.
	Included questions asking specifically about <ul style="list-style-type: none"> <li>• Romantic relationships,</li> <li>• Staying elsewhere because of a financial situation,</li> <li>• Staying elsewhere to care for someone</li> </ul>
For anyone who said they stayed at a second location, asked a series of questions about how often they stay there	For anyone who said they stayed at a second location, asked a series of questions about how often they stay there
	For anyone who said they stayed at a second location, asked who the people were who usually lived at that second location
Asked a series of questions to determine if “anyone” was in a GQ facility on April 1, such as: <ul style="list-style-type: none"> <li>• Independent or assisted living facility</li> <li>• Nursing home or skilled nursing facility</li> <li>• Correctional facility such as a jail, detention center, or prison</li> <li>• Emergency, transitional, or domestic violence shelter</li> <li>• Group home</li> <li>• Some other group facility</li> </ul>	Only asked questions about GQs if the person was counted in a GQ, and then asked specifically about the type of GQ we suspected they were counted in
For duplicates that appeared on two census questionnaires and provided the same telephone number, the phone number was called twice (once for each address) and asked the same series of questions	For duplicates that appeared on two census questionnaires and provided the same telephone number, one contact was made and the respondent was asked about both addresses at that time

Source: TCFU Training Materials for Contractors

The success of some of these changes will be discussed in Section 5.4.3.

#### **5.4.2.2 Qualitative Study**

The second sub-study was the qualitative study, which used a flexible style of interviewing that is similar to ethnographic interviewing. A semi-structured protocol provided the base, though interviewers had the freedom to probe and ask additional questions to pursue any issues or comments raised by respondents. A different protocol was used in Round 1 and Round 2. Both protocols had the same intent, but used different approaches to eliciting situations that caused duplication. In Round 2, five dependent interviews were carried out. For these cases, after receiving permission from a household in the cognitive study, a second household involved in the duplication was contacted and interviewed. The stories told by each side of these pairs will be discussed in Section 5.4.5.

Recruiting for the qualitative study was limited to HU-HU duplications with the exception of one HU-GQ case.

### **5.4.3 Findings from the Cognitive Interviewing with Suspected Duplicates**

Section 5.4.3 discusses the results obtained from the 226 cognitive interviews using the TCFU questionnaire with suspected duplicates. Table 81 first shows how the 226 interviews were distributed across the 27 categories targeted for this study. These cases were recruited purposively. Differences in rates should not be taken as differences in prevalence.

**Table 81: Distribution of Cases Cognitively Interviewed**

Description of Case	Total	
	Number	Percent
<b>Type 1 Cases – Phone Match</b>	<b>42</b>	<b>18.6</b>
1a: Whole household match: All duplicates are adults	9	4.0
1b: Whole household match: All duplicates are seniors	6	2.7
1c: Whole household match: The matches include children	5	2.2
1d: Whole household match: Any other combination of ages	4	1.8
1e: Partial household match: All duplicates are adults	7	3.1
1f: Partial household match: All duplicates are seniors	5	2.2
1g: Partial household match: All duplicates are children	1	0.4
1h: Partial household match: Other age combinations with children	4	1.8
1i: Partial household match: All other combination of ages	1	0.4
<b>Type 2 Cases – Nonphone Match</b>	<b>125</b>	<b>55.3</b>
2a: Whole household match: All duplicates are adults	17	7.5
2b: Whole household match: All duplicates are seniors	7	3.1
2c: Whole household match: The matches include children	10	4.4
2d: Whole household match: Any other combination of ages	4	1.8
2e: Partial household match: All duplicates are adults	26	11.5
2f: Partial household match: All duplicates are seniors	24	10.6
2g: Partial household match: All duplicates are children	19	8.4
2h: Partial household match: Other age combinations with children	13	5.8
2i: Partial household match: Any other combination of ages	5	2.2
<b>Type 3 Cases – HUGQ links</b>	<b>59</b>	<b>26.1</b>
3a: Military GQ	4	1.8
3b: College GQ	12	5.3
3c: Jail GQ	11	4.9
3d: Juvenile GQ	3	1.3
3e: Group Home GQ	7	3.1
3f: Nursing Home GQ	8	3.5
3g: Homeless GQ	8	3.5
3h: Workers GQ	5	2.2
3i: Religious GQ	1	0.4
<b>Total Interviews</b>	<b>226</b>	<b>100.0</b>

Source: Qualitative Interview Report

There were 226 cognitive interviews conducted; 55.3 percent were Type 2 cases, 26.1 percent were Type 3 cases, and 18.6 percent were Type 1 cases. The specific types with the most interviews were all Type 2 cases: partial household cases either with all duplicates being adults or with all duplicates being seniors each accounted for over 10 percent of all cognitive interviews.

Table 82 shows how many interviews were conducted with a person who was a suspected duplicate and how many were conducted with a proxy respondent.

**Table 82: Distribution of Cases Cognitively Interviewed, by Respondent**

Description of Case	Respondent			
	Duplicate		Proxy	
	Number	Percent	Number	Percent
<b>Type 1 Cases – Phone Match</b>	<b>37</b>	<b>28.5</b>	<b>5</b>	<b>5.2</b>
1a: Whole household match: All duplicates are adults	9	6.9	0	0.0
1b: Whole household match: All duplicates are seniors	6	4.6	0	0.0
1c: Whole household match: The matches include children	5	3.8	0	0.0
1d: Whole household match: Any other combination of ages	4	3.1	0	0.0
1e: Partial household match: All duplicates are adults	6	4.6	1	1.0
1f: Partial match: All duplicates are seniors	2	1.5	3	3.1
1g: Partial match: All duplicates are children	0	0.0	1	1.0
1h: Partial match: Other age combinations with children	4	3.1	0	0.0
1i: Partial match: All other combination of ages	1	0.8	0	0.0
<b>Type 2 Cases – Nonphone Match</b>	<b>71</b>	<b>54.6</b>	<b>54</b>	<b>56.3</b>
2a: Whole household match: All duplicates are adults	17	13.1	0	0.0
2b: Whole household match: All duplicates are seniors	6	4.6	1	1.0
2c: Whole household match: The matches include children	10	7.7	0	0.0
2d: Whole household match: Any other combination of ages	4	3.1	0	0.0
2e: Partial household match: All duplicates are adults	16	12.3	10	10.4
2f: Partial household match: All duplicates are seniors	10	7.7	14	14.6
2g: Partial household match: All duplicates are children	0	0.0	19	19.8
2h: Partial household match: Other age combinations with children	5	3.8	8	8.3
2i: Partial household match: Any other combination of ages	3	2.3	2	2.1
<b>Type 3 Cases</b>	<b>22</b>	<b>16.9</b>	<b>37</b>	<b>38.5</b>
3a: Military GQ	1	0.8	3	3.1
3b: College GQ	1	0.8	11	11.5
3c: Jail GQ	2	1.5	9	9.4
3d: Juvenile GQ	0	0.0	3	3.1
3e: Group Home GQ	5	3.8	2	2.1
3f: Nursing Home GQ	5	3.8	3	3.1
3g: Homeless GQ	5	3.8	3	3.1
3h: Workers GQ	2	1.5	3	3.1
3i: Religious GQ	1	0.8	0	0.0
<b>Total Interviews</b>	<b>130</b>	<b>100.0</b>	<b>96</b>	<b>100.0</b>

Source: Qualitative Interview Report



Of the 226 cognitive interviews conducted, 57.5<sup>46</sup> percent were conducted with an actual duplicated person, and 42.5<sup>47</sup> percent were conducted with a proxy (a household member reporting for a duplicate). Table 82 shows that this ratio varies by the type of case. Since persons under the age of 18 could not be respondents in this study, all 2g cases (Partial household match: All duplicates are children) had to be completed with a proxy by definition. It was also more difficult in some GQ cases (such as jail or college situations) to interview the duplicate. As with Table 81, researchers had some control over whether cases interviewed were duplicates or proxies, so these rates should not be taken as prevalence rates.

#### 5.4.3.1 Success Rate in Confirming the Duplication

One objective of this research was to confirm the suspected duplication. Ideally, the respondent affirms in the TCFU interview that the suspected duplicate spent time elsewhere, and then provides the address for that place. Upon review, the researcher determines that place was the same address for the other half of the link.

Table 83 shows for the three primary types of cases, how often the interviews were able to verify the duplication. Non-verified duplicates were cases where the research staff believed it was, or could have been, the same person listed on each side of the link, but the respondent did not provide enough information in the interview to confirm that. False matches, or suspected false matches, were cases that were reviewed by interviewers and Census Bureau staff and (based on the information obtained in the interview as well as the information provided on the initial census questionnaires) are thought to represent two different people on each side of the link, instead of an actual duplication. In a few cases, there may be room for disagreement about the coding of the final case outcome

---

<sup>46</sup>  $130/226=57.5$  percent.

<sup>47</sup>  $100 - 57.5 = 42.5$  percent

**Table 83: Success Rate in Cognitive Interviews, by Case Type**

Type of Case	Verified duplicate		Non-verified duplicate		False match, suspected false match, or housing unit mix-up		Total	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent
1: Phone Match	41	97.6	1	2.4	0	0.0	42	100.0
2: Nonphone Match	74	59.2	27	21.6	24	19.2	125	100.0
3: GQs	41	69.5	10	16.9	8	13.6	59	100.0
<b>Total</b>	<b>156</b>	<b>69.0</b>	<b>38</b>	<b>16.8</b>	<b>32</b>	<b>14.2</b>	<b>226</b>	<b>100.0</b>

Source: Qualitative Interview Report

All but one of the 42 Type 1 cases were successfully verified during the cognitive interview (a 97.6 percent success rate). Type 3 cases were verified 69.5 percent of the time, and Type 2 cases were verified 59.2 percent of the time (with one housing unit mix-up). One Type 2 case was considered to be a housing unit mix-up, where the same people were listed at two addresses by mistake because of either mail being delivered to the wrong address or by an interviewer using the incorrect form of the address for an enumeration.

Just under twenty percent of Type 2 cases (19.2 percent) were classified as either a false match or a suspected false match. There were 13.6 percent of Type 3 cases classified as false matches or suspected false matches. Overall, the success rate shown in Table 83 is fairly high, though the false match rate was higher than expected.

Of the 42 Type 1 cases, only one participant did not recognize the second address when both addresses were mentioned at the beginning of the interview. For that respondent, the second address was likely the address of a roommate, who moved out in 2010, but both addresses were associated with the respondent's cell phone number. The respondent was a little confused at the beginning of the interview for that case as he did not recognize the second address, but soon understood how to respond. This was the only Type 1 case where the duplicate address was not revealed by the respondent.

Seven of the Type 1 or Type 2 cases were verified during the debriefing portion of the cognitive interview, but not during the administration of the TCFU questionnaire itself.

The following three tables show the success rate for each of the nine subcategories within the Type 1, Type 2, and Type 3 cases. Table 84 first presents the results of Type 1 cases.

**Table 84: Success Rate in Cognitive Interviews, Type 1 Cases**

Detailed Case Type	Verified Duplicate		Non-verified Duplicate		False match, or suspected false match		Total	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent
1a: Whole household match: All duplicates are adults	9	100.0	0	0.0	0	0.0	9	100.0
1b: Whole household match: All duplicates are seniors	6	100.0	0	0.0	0	0.0	6	100.0
1c: Whole household match: The matches include children	5	100.0	0	0.0	0	0.0	5	100.0
1d: Whole household match: Any other combination of ages	4	100.0	0	0.0	0	0.0	4	100.0
1e: Partial household match: All duplicates are adults	6	85.7	1	14.3	0	0.0	7	100.0
1f: Partial household match: All duplicates are seniors	5	100.0	0	0.0	0	0.0	5	100.0
1g: Partial household match: All duplicates are children	1	100.0	0	0.0	0	0.0	1	100.0
1h: Partial household match: Other age combinations with children	4	100.0	0	0.0	0	0.0	4	100.0
1i: Partial household match: All other combination of ages	1	100.0	0	0.0	0	0.0	1	100.0

Detailed Case Type	Verified Duplicate		Non-verified Duplicate		False match, or suspected false match		Total	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent
<b>Type 1 Case – Phone Match</b>	<b>41</b>	<b>97.6</b>	<b>1</b>	<b>2.4</b>	<b>0</b>	<b>0.0</b>	<b>42</b>	<b>100.0</b>

Source: Qualitative Interview Report

Table 84 shows that the one Type 1 case which was not verified was a “Partial match: All duplicates adults”.

Table 85 shows the success rate for detailed Type 2 cases.

**Table 85: Success Rate in Cognitive Interviews, Type 2 Cases**

Detailed Case Type	Verified Duplicate		Non-verified Duplicate		False match, or suspected false match		Total	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent
2a: Whole household match: All duplicates are adults	14	82.4	2	11.8	1	5.9	17	100.0
2b: Whole household match: All duplicates are seniors	4	57.1	1	14.3	2	28.6	7	100.0
2c: Whole household match: The matches include children	7	70.0	3	30.0	0	0.0	10	100.0
2d: Whole household match: Any other combination of ages	2	50.0	2	50.0	0	0.0	4	100.0
2e: Partial household match: All duplicates are adults	12	46.2	7	26.9	7	26.9	26	100.0
2f: Partial household match: All duplicates are seniors	11	45.8	4	16.7	9	37.5	24	100.0
2g: Partial household match: All duplicates are children	14	73.7	1	5.3	4	21.1	19	100.0
2h: Partial household match: other age combinations with children	9	69.2	4	30.8	0	0.0	13	100.0
2i: Partial household match: any other combination of ages	1	20.0	3	60.0	1	20.0	5	100.0
<b>Type 2 Case – Nonphone Match</b>	<b>74</b>	<b>59.2</b>	<b>27</b>	<b>21.6</b>	<b>24</b>	<b>19.2</b>	<b>125</b>	<b>100.0</b>

Source: Qualitative Interview Report

For Type 2 cases, subsets 2e and 2f (partial match cases with adults and seniors, respectively) had less than half of the cases verified as being actual duplicates. The one false match case in subset 2c (Whole match: the matches included children) was the housing unit mix-up described after Table 83.

Table 86 shows the success rate for detailed Type 3 cases.

**Table 86: Success Rate in Cognitive Interviews, Type 3 Cases**

Detailed Case Type	Verified Duplicate		Non-verified Duplicate		False match, or suspected false match		Total	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent
3a: Military GQ	3	75.0	1	25.0	0	0.0	4	100.0
3b: College GQ	9	75.0	0	0.0	3	25.0	12	100.0
3c: Jail GQ	9	81.8	1	9.1	1	9.1	11	100.0
3d: Juvenile GQ	1	33.3	1	33.3	1	33.3	3	100.0
3e: Group Home GQ	1	14.3	4	57.1	2	28.6	7	100.0
3f: Nursing Home GQ	8	100.0	0	0.0	0	0.0	8	100.0
3g: Homeless GQ	6	75.0	2	25.0	0	0.0	8	100.0
3h: Workers GQ	4	80.0	1	20.0	0	0.0	5	100.0
3i: Religious GQ	0	0.0	0	0.0	1	100.0	1	100.0
<b>Type 3 Case</b>	<b>41</b>	<b>69.5</b>	<b>10</b>	<b>16.9</b>	<b>8</b>	<b>13.6</b>	<b>59</b>	<b>100.0</b>

Source: Qualitative Interview Report

Table 86 shows that one of the three links to a Juvenile GQ was determined to be a false match and three of the twelve links with a college GQ were determined to be a false match. All eight nursing home cases were verified. Links to group home and juvenile GQs had the lowest rates of verification. For group homes, it seems as though these were often a case of mistaken GQ identification; in reviewing the cases and addresses after the interview, it appeared as though some of these group homes might actually have been correctional facilities. Since the TCFU instrument specifically asked about group homes (including residential treatment facilities, residential schools for people with disabilities, and such), and then vaguely asked about any other places where the duplicate might stay, this targeted approach might have been too narrow for these cases. Four of the five workers' GQ cases were technically verified but the workers' GQ category also contained some misclassification, as three cases involved duplication to boarding schools instead of a true workers' GQ. Boarding school students should not actually be enumerated in the census at the boarding school if they have a usual home elsewhere.

Of the 155 interviews where the duplication was verified, Table 87 shows a simplified list of reasons that were considered to be the cause of the duplication.

**Table 87: For Verified Duplications, Reasons for Duplication by Type of Case**

Reason for Duplication	Type 1 Cases		Type 2 cases		Type 3 cases		TOTAL	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent
Completed wrong form	2	4.9	0	0.0	0	0.0	2	1.3
Convenience address <sup>48</sup>	1	2.4	6	8.2	0	0.0	7	4.5
Custody	0	0.0	6	8.2	0	0.0	6	3.9
GQ	0	0.0	0	0.0	41	100.0	41	26.5
Moved	18	43.9	31	42.5	0	0.0	49	31.6
Other Property	16	39.0	14	19.2	0	0.0	30	19.4
Relatives' Household	4	9.8	16	21.9	0	0.0	20	12.9
<b>Total verified cases</b>	<b>41</b>	<b>100.0</b>	<b>73</b>	<b>100.0</b>	<b>41</b>	<b>100.0</b>	<b>155</b>	<b>100.0</b>

Source: Qualitative Interview Report

The most common reason for duplication among all the verified cases was a move. Moving and having other property were the two main reasons for duplication among Type 1 cases. Those were also two of the three most common reasons for duplication among the Type 2 cases, in addition to staying at a relative's house.

### Success Rate by Type of Respondent

Table 82 showed that 130 interviews were conducted with persons who were themselves suspected duplicates, while 96 interviews were conducted with proxy respondents. Table 88 shows the success rate of the cognitive interviews by respondent type.

**Table 88: Success Rate in Cognitive Interviews, by Respondent**

Case Outcome	Duplicate		Proxy	
	Number	Percent	Number	Percent
Verified duplicate	90	69.2	65	67.7
Non-verified duplicate	19	14.6	19	19.8
Housing unit mix-up	1	0.8	0	0.0
False match	10	7.7	4	4.2
Suspected false match	10	7.7	8	8.3
<b>Total</b>	<b>130</b>	<b>100.0</b>	<b>96</b>	<b>100.0</b>

Source: Qualitative Interview Report

Table 88 shows that interviews with suspected duplicates led to the duplication being verified 69.2 percent of the time, and interviews with proxy respondents led to the duplication being verified 67.7 percent of the time. The interviews with suspected duplicates were classified as false matches 7.7 percent of the time, but only 4.2 percent of the time for interviews with proxies. Since the false match designation was assigned after reviewing the data and conducting the interview, it was easier to classify a case as a false match if we actually spoke with the duplicate and they sincerely gave no indication in the interview of having any other place to stay

<sup>48</sup> A convenience address is an address used for mail delivery and legal documents such as licenses, rather than for habitation.

that resembled what we expected. Proxy cases were classified as non-verified duplicates 19.8 percent of the time, compared to only 14.6 percent of the time in interviews with actual duplicates. This could be a result of either a true duplication that the proxy just could not provide enough information to us about, or an actual false match duplication that we were reluctant to classify as such based only on the proxy's interview responses.

### Success Rate by Number of Duplicated Persons

Links with more than one person identified as a suspected duplicate are matched with greater confidence than links with just one duplicate, since multiple person duplicates provide more information on which to match. The following section shows how often interviews in the cognitive study led to verified duplication based on the number of persons linked. The 59 Type 3 cases are not analyzed in this section because they could only contain one duplicated person in the link.

Table 85 shows the success rate between interviews with respondents from households that had multiple duplicates compared to interviews with respondents from households that had only one duplicate.

**Table 89: Success Rate in Cognitive Interviews, by Number of Duplicates**

Case Outcome	One-person Duplicates		Multiple Duplicates	
	Number	Percent	Number	Percent
Verified duplicate	55	61.1	59	76.6
Non-verified duplicate	12	13.3	16	20.8
Housing unit mix-up	0	0.0	1	1.3
False match	10	11.1	0	0.0
Suspected false match	13	14.4	1	1.3
<b>Total</b>	<b>90</b>	<b>100.0</b>	<b>77</b>	<b>100.0</b>

Source: Qualitative Interview Report

NOTE: These numbers are only reported for Type 1 and Type 2 cases, because Type 3 cases only had one duplicate by definition.

Table 85 suggests that interviews with respondents from households with multiple duplicates were more successful than interviews with respondents from households with only one duplicate. For multiple duplicate interviews, 76.6 percent confirmed the duplicate address, compared to 61.1 percent of one-person duplicate interviews. Interviews with one-person duplicate households also resulted in a higher rate of confirmed and suspected false matches. Only one case with multiple duplicates was a suspected false match while 23 cases (25.6 percent) with one duplicate were either confirmed or suspected false matches. As discussed in Section 5.1.5 earlier, the number of matched people in a household is a factor in assigning the probability of a match to a household during the matching algorithm (and was also a factor in the Census Bureau's ad hoc data review during this particular study). The more people who can be matched, the more likely it is that the match reflects true duplicated persons.



The cases were also categorized by being either a partial household match or a whole household match. Table 90 shows the success rate by that categorization.

**Table 90: Success Rate in Cognitive Interviews, by Household Match**

Case Outcome	Partial Households		Whole Households	
	Number	Percent	Number	Percent
Verified duplicate	64	61.0	50	80.6
Non-verified duplicate	20	19.0	8	12.9
Housing unit mix-up	0	0.0	1	1.6
False match	9	8.6	1	1.6
Suspected false match	12	11.4	2	3.2
<b>Total</b>	<b>105</b>	<b>100.0</b>	<b>62</b>	<b>100.0</b>

Source: Qualitative Interview Report

NOTE: These numbers are only reported for Type 1 and Type 2 cases.

There were 105 interviews completed with partial household duplication cases, and 62 interviews completed with whole household matches. Over eighty percent (80.6 percent) of the whole household matches were able to be verified while only 61.0 percent of the partial households were verified. False matches occurred more often in partial household matches (8.6 percent of the time) compared to whole household matches (1.6 percent).

The categorization of whole or partial household matches can be further delineated by the actual number of duplicates within those matches. This is shown in Table 91.

**Table 91: Success Rate in Cognitive Interviews, by Household Match and Number of Duplicates**

Case Outcome	Partial Households				Whole Households			
	One Duplicate		Multiple Duplicates		One Duplicate		Multiple Duplicates	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent
Verified duplicate	44	57.9	20	69.0	11	78.6	39	81.3
Non-verified duplicate	12	15.8	8	27.6	0	0.0	8	16.7
Housing unit mix-up	0	0.0	0	0.0	0	0.0	1	2.1
False match	9	11.8	0	0.0	1	7.1	0	0.0
Suspected false match	11	14.5	1	3.4	2	14.3	0	0.0
<b>Total</b>	<b>76</b>	<b>100.0</b>	<b>29</b>	<b>100.0</b>	<b>14</b>	<b>100.0</b>	<b>48</b>	<b>100.0</b>

Source: Qualitative Interview Report

NOTE: These numbers are only reported for Type 1 and Type 2 cases.

Table 91 shows that partial household matches with only one duplicated person had the fewest percent of cases verified (57.9 percent) and the highest percent of false matches. In making these decisions after interviews, we looked at relationship and household composition, which are

not included in the current computer-matching process. This suggests researching a way to include relationship or household composition in the computer-matching process.

#### 5.4.3.2 Address Completeness

To make a decision about the one place where a duplicated person should be counted, three pieces of information were needed from the followup interview: the address of the other location, the dates the person stayed at the other address, and how often they stay at that other address. This allows the Census Bureau to determine where each person should have been counted in the census.

The completeness of the address information was essential to verify the duplication and, in the future, would be critical to actually removing an individual from one roster. For this study, interviewers recorded if the address data they received were complete (meaning that at least the street number, street name, city, and state were provided) or partial. The frequency of complete address data might be distorted in this study since some respondents referred to address books they had on hand during the face-to-face interviews where they were receiving an incentive, which they may not do if they are responding via a non-incentivized telephone interview.

Table 92 shows how often the collected addresses were considered to be partial or complete. This table contains all addresses reported by respondents, not just those for verified duplication cases.

**Table 92: Frequency of Address Completeness**

<b>Completeness of Addresses Provided</b>	<b>Number</b>	<b>Percent</b>
Complete	242	59.3
Partial	166	40.7
<b>Total Addresses Collected</b>	<b>408</b>	<b>100.0</b>

Source: Qualitative Interview Report

Some individuals provided more than one address during the interview, often including addresses that were not related to the duplication, so a total of 408 addresses were recorded in the 226 interviews. Almost sixty percent (59.3 percent) were considered to be complete addresses.

Table 93 shows how often addresses were collected for specific probes within the TCFU interview, and shows the frequency of complete addresses compared with partial addresses for each probe.

**Table 93: Number of Complete and Partial Addresses Provided by TCFU Questions**

Type of Question	Partial Addresses		Complete Addresses		Total Addresses	
	Number	Percent	Number	Percent	Number	Percent
Did [duplicate] move?	11	9.5	105	90.5	<b>116</b>	<b>100.0</b>
Stay with another parent (Child duplicates only)	20	47.6	22	52.4	<b>42</b>	<b>100.0</b>
Stay with a parent (Adults and seniors)	12	46.2	14	53.8	<b>26</b>	<b>100.0</b>
Stay with a son/daughter	12	57.1	9	42.9	<b>21</b>	<b>100.0</b>
Own seasonal/vacation home	5	27.8	13	72.2	<b>18</b>	<b>100.0</b>
Current address (Section C)	2	11.8	15	88.2	<b>17</b>	<b>100.0</b>
Own another residence	0	0.0	17	100.0	<b>17</b>	<b>100.0</b>
Stay with a significant other	11	78.6	3	21.4	<b>14</b>	<b>100.0</b>
Nursing home GQ	12	85.7	2	14.3	<b>14</b>	<b>100.0</b>
Stay with another relative	11	84.6	2	15.4	<b>13</b>	<b>100.0</b>
Laundry list questions <sup>49</sup>	8	66.7	4	33.3	<b>12</b>	<b>100.0</b>
College GQ	10	83.3	2	16.7	<b>12</b>	<b>100.0</b>
Stay at a nursing home	4	36.4	7	63.6	<b>11</b>	<b>100.0</b>
Jail GQ	6	60.0	4	40.0	<b>10</b>	<b>100.0</b>
Stay at any other GQ	5	55.6	4	44.4	<b>9</b>	<b>100.0</b>
Stay somewhere for health problems	9	100.0	0	0.0	<b>9</b>	<b>100.0</b>
Stay with a brother/sister	6	75.0	2	25.0	<b>8</b>	<b>100.0</b>
Stay at a nursing home, jail, emergency shelters	2	33.3	4	66.7	<b>6</b>	<b>100.0</b>
Homeless GQ	3	60.0	2	40.0	<b>5</b>	<b>100.0</b>
Military GQ	4	80.0	1	20.0	<b>5</b>	<b>100.0</b>
Another address for school	0	0.0	4	100.0	<b>4</b>	<b>100.0</b>
Juvenile GQ	3	75.0	1	25.0	<b>4</b>	<b>100.0</b>
Workers GQ	1	33.3	2	66.7	<b>3</b>	<b>100.0</b>
Stay at a college, university, or graduate school	1	50.0	1	50.0	<b>2</b>	<b>100.0</b>
Stay somewhere for a job	0	0.0	2	100.0	<b>2</b>	<b>100.0</b>
Group home GQ	2	100.0	0	0.0	<b>2</b>	<b>100.0</b>
Stay with a friend, neighbor, or legal guardian	2	100.0	0	0.0	<b>2</b>	<b>100.0</b>
Stay at a rehabilitation hospital	2	100.0	0	0.0	<b>2</b>	<b>100.0</b>
Stay at another residence	1	100.0	0	0.0	<b>1</b>	<b>100.0</b>
Stay somewhere for child care	1	100.0	0	0.0	<b>1</b>	<b>100.0</b>
<b>Total</b>	<b>166</b>	<b>40.7</b>	<b>242</b>	<b>59.3</b>	<b>408</b>	<b>100.0</b>

Source: Qualitative Interview Report

<sup>49</sup> The laundry list of situations was asked at the end of the interview as a final attempt to probe the respondent on a number of unconventional situations that may have led to duplication. Some situations included in this list were training or internships, athletics or sports, to care for someone, because of financial problems, or because of someone's job.

The question about moving generated the most addresses (116 were collected) and the majority of those addresses (related to a move) were considered to be complete addresses (90.5 percent). If the respondent reported owning another property, they provided a complete address one hundred percent of the time. Partial addresses were prevalent when the duplicate was reported to have stayed with a significant other (78.6 percent of such addresses were partial), son or daughter (57.1 percent partial), brother or sister (75.0 percent partial), or another relative (84.6 percent partial). Partial addresses were also common when reporting on GQ stays. Only workers GQ cases had more complete addresses than partial addresses provided.

The only question in the TCFU interview that did not result in any addresses being collected was the question to adult duplicates asking if they stayed with friends.

### **5.4.3.3 TCFU Question Comprehension**

Respondents in this study were debriefed after the TCFU instrument was administered to understand how they had perceived the questions in the instrument. During that debriefing, respondents reported having some confusion or uncertainty about the meaning of select TCFU questions.

The meaning of the word “stay” provoked the biggest confusion. Many participants asked for a definition of “stay,” and some felt they could not answer the question since no definition was provided. For example, one respondent was unsure if “stay” meant “reside” or “visit.” Another respondent commented that the definition of “stay” to them meant living somewhere and would involve being somewhere for more than a month. Generally, “staying with someone” was a term that needed to be defined. Many respondents asked for a definition or volunteered one when answering the question.

A few people also were confused when answering for just 2010 (given most interviews took place in 2011). In addition, some participants were not answering for the whole year, but rather from April 1 until the end of December 2010. For example, one respondent said she had not moved because she was thinking about whether she had moved from her current residence, but she later revealed that she moved into her current residence in March 2010.

The TCFU instrument repeatedly referenced April 1, 2010, since that is considered to be ‘Census Day’ and an important component of assigning residency status. The use of this specific date provoked some negativism; participants wanted to know why that particular date was of interest. One respondent stated: “The question that you kept asking about April 1st made me wonder what they are after. It was slightly creepy.”

The questions about how often a person stayed somewhere also brought up confusion during the interview. Responses often had explanations such as “half the time, because I go to work all week.” One respondent indicated staying at home “less than half the time” and explaining: “I only stayed there after hours when I’m not at work and school.” The question was not intended to get how often a person was physically in the housing unit as opposed to being at work or

school or otherwise away from the physical structure. Rather, the question was intended to solicit how people divided their time between housing units, similar to understanding a custody arrangement two parents might have for their child.

Generally, the TCFU questions were understood as intended. However, sometimes, knowing the respondent's situation, it was obvious he/she was not providing an accurate response. Moving situations yielded the most incorrect responses,<sup>50</sup> especially when there were multiple duplicates that moved together (e.g., parents and their children), or a proxy moved with the duplicate. The phrase "other than the places you already mentioned" used in most gate questions seemed to induce confusion as respondents often misinterpreted it while reporting on additional family members. For instance, a parent would report having moved, but then when the question was asked in regards to a child, they would not say that the child had moved. The question at that point was misinterpreted by the respondent; they heard it as asking if the duplicated child had moved away from the parent.

#### **5.4.3.4 Privacy and Confidentiality Concerns**

A total of 55 respondents (of the 226 interviewed) voiced concerns about privacy during the interview. Among those, 26 were proxy respondents and 29 were actual duplicated persons. Asking for other people's addresses and asking whether the duplicated person stayed with a girlfriend or a boyfriend provoked the most comments.

Asking parents about their children also elicited concerns, especially when the child was the only duplicate in the household. Respondents were not informed in advance about the details of the survey (e.g., about whom the survey would be asking), and many assumed the interview would be about themselves. A few respondents mentioned that they would not have agreed to do the interview if they knew the questions would focus only on their son or daughter.

Asking respondents to provide names of people who usually live at a particular address was found to be alarming. This question was a new part of the TCFU; it was thought to potentially be useful in confirming the duplicate address if there was a name match between the TCFU information and the household roster of the suspected duplicate address. However, asking respondents to provide names associated with addresses or other people's addresses elicited the most refusals.

No TCFU questions were coded as provoking confidentiality concerns. Only seven of the 226 participants expressed concerns related to the use of their data or were worried about the ulterior motive of the TCFU questionnaire. For example, one respondent insisted on clarifying that the information she gave would not be going anywhere further, making sure her information would not be saved or reserved. She said she came "from an older era where there is a lot of suspicion, so that's why I'm curious." Similarly, another respondent requested the audiotape from the interview to be mailed back to him.

---

<sup>50</sup> This is probably not surprising given movers were the largest category among cases with uncovered duplicate addresses (32 percent of the time).

Some of the targeted GQ questions (staying at a shelter, a jail, etc.) were perceived as sensitive by respondents and led to untruthful responses. For example, one participant denied staying at a shelter during the TCFU, but admitted to it during the cognitive portion of the interview. He simply stated he did not want to talk about it.

The total number of refusal codes in this study was very low however. This might be a result of the face-to-face mode of data collection and the monetary incentive for participation. Respondents were also shown the interviewer's Census Bureau badges to verify the legitimacy of the study. During the cognitive and debriefing portion of the interviews, several respondents expressed hesitation or apprehension when asked "would your responses be any different if we had called on the phone..." and "if we were on the phone and I was going through the introduction, do you think it would have affected your decision to participate." Those who were explicit about their concern cited fraud, privacy, and not being able to confirm or be sure who is on the other end of the phone. Thus, this study likely provides a very conservative idea of respondents' unwillingness to reveal the information of interest.

#### **5.4.3.5 Assessing the Changes from the CFU Instrument**

As described in Section 5.4.2.1, the TCFU instrument was created to ask targeted questions about the duplicated persons based on

- whether the match was a household-household match (Type 1 and Type 2 cases) or household-group quarters (GQ) match (Type 3 cases);
- Whether it was a phone number match case (Type 1 compared with Type 2 cases); and
- the age of the duplicated persons.

Overall, this strategy seemed to work well and was adequate for various living situations.

The TCFU instrument was designed to provide both addresses to respondents when a phone match existed. This strategy seemed to work well, and only one participant did not recognize the second address. This was the only Type 1 case where the duplicate address was not revealed. In all other Type 1 cases, presenting both addresses was perceived well by respondents and no one expressed any privacy concerns.

A small number of detailed questions were found to be age inappropriate. Some respondents laughed when we asked about hobbies and activities of people 70 years of age or older, or when asked whether they were in the military in 2010.

The GQ-specific probes for Type 3 cases verified the duplicate in 69.5 percent of the cases and did not reveal the duplication in 16.9 percent of cases. Of the ten Type 3 cases where the duplication was not revealed, four were duplicates to a group home facility. Two of those might have been misclassified; during review of the data, the addresses appeared to be correctional facilities rather than group homes. The relevant situations to reveal the duplication might not have come up during the tailored interview then. Misclassifications might be the primary barrier to successfully using a GQ-specific targeted followup in the future.

### 5.4.4 Findings from the Qualitative Interviewing with Suspected Duplicates

As discussed at the beginning of Section 5.4, the qualitative research had two components; the cognitive interviews (described in the previous section) and the qualitative interviews. This section will discuss the result of the 50 qualitative interviews conducted with suspected duplicate persons.

The main aim of the qualitative study was to shed light on the processes and events within households that might have resulted in duplication. These interviews were semi-structured ethnographic-style and had a primary objective of providing background insight into the social and procedural processes that may result in duplication in the census.

#### 5.4.4.1 Success Rate in Confirming the Duplication

As a measure of success, Table 94 shows the frequency with which the qualitative interviews verified the duplication within the interview.

**Table 94: Success Rate in Qualitative Interviews**

<b>Case Outcome</b>	<b>Number</b>	<b>Percent</b>
Verified duplication	35	70.0
Non-verified duplication	11	22.0
False match	4	8.0
<b>Total</b>	<b>50</b>	<b>100.0</b>

Source: Qualitative Interview report

Of the 50 qualitative interviews, 35 (70.0 percent) verified the duplication within the interview.

Cases that were considered "verified" were those in which the interviewer was able to collect information that helped to explain the duplicate. In making this assessment, information provided by respondents was compared with information provided by the Census Bureau about the names, relationships, and addresses of duplicated persons. In assessing success, a strict standard was not followed of a perfect address-to-name match. Instead, partial address information provided by the respondent, relationship data, and other cues were used to arrive at these conclusions. Cases listed as false matches were those in which nothing in the interview suggested a connection to the second household, the names or ages in the links did not match perfectly, or relationships indicated that it was unlikely to be the same individual (e.g., if a child was listed as a biological child of two seemingly different fathers). In a few cases, there may be room for disagreement about the coding of the final case outcome.

Twenty-five interviews were conducted in fall 2010 (considered Round 1). Special emphasis was placed on recruiting cases involving transitioning seniors,<sup>51</sup> young adults aged 18 to 29 and

---

<sup>51</sup> Transitioning seniors are adults 50 to 80 years old who may be duplicated due to age-related reasons such as a move related to downsizing or into an assisted living facility. Also included are seniors who may have multiple

families with children. In Round 2 (spring and summer 2011), an effort was made to screen out households in which the duplication stemmed from moving from one address to another, since those had been the bulk of Round 1 duplications. A question was inserted into the recruiting script to deal with this situation. Recruiters were instructed to focus recruiting on children duplicated because of potential custody situations and cases that might be second homes.

Table 95 examines the reasons for duplication seen in the 35 interviews with the duplication verified within the interview. Again, these interviews were purposively recruited, so rates represented in Table 95 should not be seen as prevalence rates.

**Table 95: Reasons for Duplication in Qualitative Interviews**

<b>Reason for duplication</b>	<b>Number</b>	<b>Percent</b>
Moved	10	28.6
Custody situation	13	37.1
GQ stay (health, rehabilitation)	2	5.7
Convenience address	3	8.6
High Mobility	1	2.9
Second home	6	17.1
<b>Total</b>	<b>35</b>	<b>100.0</b>

Source: Qualitative Interview report

The most frequent reason for duplication was because of a custody situation, with movers as the second most common. One case labeled as due to a GQ stay involved an enumeration at an apartment facility where rehabilitation was provided. Although it was technically a housing unit, the reason for the duplication was related to the provision of health services to an elderly respondent on a temporary basis. It is categorized here as a "GQ" because of its underlying cause.

The category of "convenience address" was used for situations in which a duplication was created when a person lived primarily at one address but used a second address to receive mail, or for legal purposes such as a driver's license.

The category "highly mobile person" describes individuals who travel between stable homes of friends and/or relatives. In this instance, the duplicated person was enumerated at both his mother's house and his girlfriend's house.

#### **5.4.4.2 Complex Ties: Presence at an Address or Connection to an Address**

In a number of situations, complex ties between interconnected households affected the duplications. This was most evident in custody cases since not all situations where children have a presence in two households are covered by formal agreements such as divorce decrees. Especially with grandparents and aunts or uncles, the custody may be flexible, inconsistent, and

---

residences in order to take advantage of temperate climates (snow birds) or who live with or move between their adult children's homes.



part of a set of difficult relationships. We have termed this "informal custody." These situations may not be reported as "custody" in questions that use this term since relatives describe the child's presence in their household in other ways. In addition, parents who have not formally transferred custody may report the child's presence, regardless of the amount of time spent in their home. Complex family ties also affected the reporting of second homes. Such ties may reflect economic ties between related families—families that co-own property or use properties for other economic reasons, such as family-based businesses, may report members in both places regardless of the actual amount of time spent there. Convenience addresses (where a person uses an address to receive mail or for other legal purposes) also appeared generally to reflect family ties—we learned that people maintain an address with a particular relative, and may be reported there even when they are not there most of the time.

Respondents seemed to err by including family members without reference to the actual amount of time spent in the household. It is likely that they believed these reports were in some sense accurate, since they described the family unit. To improve reporting, the census form must stress the essential core concept: the actual presence of an individual – the “most of the time” concept, not the familial or economic tie.

Such complex situations may also affect response by effecting interpretation of survey cues. We observed that the simple cues for duplication situations did not always work to elicit these stories (e.g., asking if a child stays with another parent did not always generate the response), and other cues than those that were intended were used (e.g., when a custody situation was described as a "second home"). This may reflect the complexity of the situations. Thus, we concluded that question strategies should include multiple ways of identifying situations where people might be duplicated. Simple screeners may not be effective.

Movers accounted for a large proportion of duplications. The duplications essentially occurred because the household was enumerated in both places. Respondents were willing to provide information about move addresses and dates. While moving is considered a complex living situation in the CFU interview and for the purposes of identifying those persons who might be difficult to enumerate, duplicated movers do not seem to be as difficult to resolve as other duplicated persons.

#### **5.4.4.3 Concept Matching**

Census Bureau definitions and concepts may affect the ability to respond "accurately" if not clearly stated. In the qualitative data, respondents' ideas did not match census definitions in the following important ways:

1. Respondents had a social rather than physical definition of the household: they were concerned with telling us who "belonged" in their household, based on other criteria than physical presence.
2. “Custody” referred only to agreements about custody, and was not extended to informal situations. In addition, parents with custody agreements granting them with a smaller amount of time may still regard this as "custody." This may lead to the erroneous enumeration of children where they spend a minor part of time.

3. Respondents believed that a person could live in two places, and did not see the necessity for identifying one place as the "right" place to enumerate someone. The idea that a person could only be counted in one place did not seem natural, and even when provided with this information, respondents were not sure why it was necessary.

If respondents are to report according to specific definitions, the Census Bureau has to provide them with information on the concepts required. To promote adequate probing, enumerators also need to understand the census definition well.

#### **5.4.4.4 Accuracy Concerns**

The protocols were designed to elicit information from the respondent without the interviewer directly mentioning the particular duplication of interest to us. This indirectness may have caused some respondents to have the sense that we were checking the data they originally submitted to the census. Respondents often indicated that they wanted to be "accurate" and showed some anxiety about a possible error on their part. In future followup interviews, we may need to provide specific assurances that we are checking up on census procedures; thus, any error may have come from a source other than their initial report. However, stressing accuracy in any way may have negative effects, too. It might easily lead to the respondent feeling the need to create a seamless story to assert the accuracy of the original report, rather than encouraging efforts to improve memory for complex situations. Respondents' concern for accuracy may in fact lessen the motivation to work on specific memory of where the person being discussed was during the reference period. A balance needs to be found between encouraging accuracy and creating anxiety about errors in reporting. Explanations provided in the interview, the recruitment script, and the general context of the interview all play a role in achieving this balance.

#### **5.4.4.5 Quality of Address Information**

Although respondents seemed to want to report accurately, in many cases they did not know the complete address of the places on which they were reporting. They offered neighborhood information, cross streets, or driving directions as a substitute. This occurred even when ex-spouses reported on the addresses where children in joint custody spent much of their time. Although this information was often adequate to verify duplication in this study and for learning about circumstances surrounding duplication, it may not be sufficient for automation of the verification procedures.

#### **5.4.5 Dependent Interviews**

The researchers attempted to interview both sides of some HU to HU links and compare the answers. These interviews with both sides of a link were referred to as dependent interviews. Dependent interviews were difficult to accomplish since both sides of a link had to be in one of the five select geographic regions where interviewing was conducted and a new willing respondent had to be contacted for the second interview, who was different from the first respondent and ideally unaware of the first interview. The universe of cases that were available for dependent interviews was restricted to Type 2 cases to meet these requirements (because

there was only one phone number available to reach Type 1 and Type 3 cases, which could only yield respondents at a single household).

Seven dependent interviews were successfully completed within this study. Often, considerable differences emerged between the versions of the duplicate's residence pattern in each side of the paired interviews. A notable pattern was the creation of a complete, self-consistent story about the presence of the duplicated person in the household within a single interview. In these instances, the respondent always told us that the duplicate lived with them. Such interviews indicate that different households have vastly different views of the membership of the individual (i.e., they each "claimed" the relative). The gap indicates that one or both accounts are inaccurate. Often, the "truth" could not be detected through the interviews.

The dependent interviews are described below for direct comparisons of the stories provided by each respondent. These situations describe two distinct interviews, with two distinct respondents, one for each side of a duplicate link.

### **First situation:**

**Cognitive Interview:** The child's mother indicated that the daughter "moved back" to live with her in March 2010 and lived at that address for the remainder of the year. This respondent appeared to be rather emotional over the question of custody as the result of some undefined "incident."

**Dependent Qualitative Interview:** The interview was with a father who described the teenage daughter as living with him and spending holidays and vacations with her biological mother. This was said to be a court-agreed custody arrangement.

### **Second situation:**

**Cognitive Interview:** The mother reported the daughter as living with her, but spent the summer with her grandmother.

**Dependent Qualitative Interview:** A grandmother reported having "guardianship" of her granddaughter, and said that the child travelled to another state to visit her mother for 1 or 2 weekends a year.

### **Third situation:**

**Cognitive Interview:** A grandmother listed her daughter and her daughter's two children as living with her for all of 2010. She indicated that they had very recently moved to a new home in May 2011. The children are not reported as visiting their father.

**Dependent Cognitive Interview:** The child's mother reported that the children lived with her in 2010 and occasionally visited their grandmother at the other address. She did not report herself as having any presence in the grandmother's household. In addition, she reported that the children visited their father at a third address.

### **Fourth situation:**

**Cognitive Interview:**

The duplicate was the respondent in the original cognitive interview, which was a Type 3 (HU to GQ) case. The two enumerations happened at a halfway house (which had been classified as a HU) and a rehabilitation facility (the GQ). The halfway house was a private housing unit for women with systematic connections to a drug rehabilitation facility. The respondent reported helping the manager at the rehabilitation facility to fill out a census form but reported being a resident of the halfway house on April 1, 2010.

**Dependent Qualitative Interview:**

The respondent in the qualitative interview lived at the halfway house, and had been the respondent for the census enumerator, providing information on everyone in the house at that time, although she was not on the census roster for that unit. The halfway house was essentially a private apartment house with upstairs and downstairs units, and the qualitative respondent regarded it as an "independent living" stage of recovery. Some of the residents in the halfway house continue to get their mail at the former rehabilitation facility and the respondent indicated that they often used the address of the rehabilitation facility as their "permanent address." The respondent in this dependent interview agreed that the respondent was at the halfway house on April 1, 2010.

**Fifth situation:****Cognitive Interview:**

The respondent in the cognitive interview was the daughter of the duplicated woman. She described her mother's presence in the other household (with a second daughter) as temporary and reported that her mother mostly lived at her address, using the phrase "permanent address". She described her mother as having gone to the other daughter's home for a few months in late 2010 (mid September to December.) Other than that, the mother was reported as going to the other daughter for weekends to "get away."

**Dependent Qualitative Interview:**

The qualitative respondent was the other daughter; she reported that her mother "was in my home at the time and had been there for a while," and is "at my house every day." This respondent knew that her sister had also listed their mother, saying the following:

*Respondent: [the census roster] is correct, but ... [my mother] was living between my house and my sister's house. I later learned that my sister included her on the census form for her address.*

*Interviewer: Could your sister have thought your mother lived with her?*

*R: No, because she was at my home at the time and she had been there for a while. But her mail goes to my sister's address and so my sister felt she should count her as a member of her household.*

*I: Does she spend time at your sister's?*

*R: No, she's at my house every day but her mail goes there. Her social security does go there. She's always used that as a permanent address. I don't know her reasoning.*

The two sides agree on the location of the mother's "permanent address," (i.e., where she gets her mail) but disagree about where she spent most of her time and where she was around the time of the census. This case appears to follow the pattern for the dependent custody cases, in that both sides "claim" the membership of the duplicated relative.

### **Sixth situation:**

#### **Cognitive Interview:**

The first interview was with the duplicate, who reported moving.

#### **Dependent Cognitive Interview:**

This interview was with the duplicate's sibling, who also reported that the duplicated person had moved.

### **Seventh situation:**

#### **Cognitive Interview:**

A woman and her two children were duplicated in this case. This first interview was conducted with the woman's mother, who reported that her daughter and the two children lived with her in 2010 and moved to a new place in May 2011. The interviewer noted that the respondent made a comment that they were very private people, and this may be part of the reason why the duplicate address was not revealed.

#### **Dependent Cognitive Interview:**

This interview was with the adult duplicate, who mentioned that the children spent "a couple of nights out of the year" at her mother's address but never confirmed that she spent time there. This respondent also revealed the father's address where the children spent time "a couple of weekends during the year." No specific dates were provided, as it was described that such stays occurred "off and on" throughout the year. Both respondents claimed the duplicated children stayed at each address the entire year. Privacy concerns and confusion about dates (such as misremembering the move) might have been the reason for the conflicting reports.

Within the dependent interviews described above, the cases are "verified" in the sense that the address of the other place is revealed in the interviews on both sides, and the identity of the duplicate is agreed upon. Beyond that, however, great differences emerge in the account of the circumstances surrounding the time spent in the two places, despite the apparent openness of the respondents on their own in these situations. Relying on both of these reports jointly, it would not be possible to determine the proper enumeration for the duplicated person in six of the seven situations. Each account on its own would imply that the proper enumeration had been established.

### **5.4.6 Memory Issues for Respondents**

Because this research was done a year to a year and a half after the census, the issue of respondent memory is important in assessing the results. In the qualitative study, we were able to note some memory issues for respondents, who became uncertain of dates or the circumstances of the initial enumerations during the census. However, in reporting the places that household members go or assessing where they generally live, we did not notice this pattern. The stories remain essentially stable, despite repeated probing. This may be related to the creation of self-consistent accounts, since the repeated probing probably would have led to at least some inconsistency in the reports.

### **5.4.7 Privacy and Sensitivity**

Respondents in both the qualitative and cognitive studies suggested that the subject of the interview itself was somewhat intrusive—questions about one’s living situation and presence of household members were perceived as “personal.” Respondents recalled census advertising messages about “counting people” and community benefits. These purposes were seen in a positive light. However, respondents do not see the connection between them and a fine-grained look at who goes where, when, and for how long.

In the qualitative interviews, there was little open complaint or refusal to provide information, but some issues appeared difficult to discuss (e.g., foreclosure, loss of jobs, and some custody situations). It seems likely that repeated probing about the same situation is necessary to elicit the full story.

A related pattern is suppression of data—the duplication is not mentioned at all. The qualitative interviews show evidence, although some of it is circumstantial, that there was information that respondents did not reveal—relationships with unmarried partners, or certain custody situations. We inferred this information from relationship and other data in census records (such as the duplicate being listed with an unmarried partner, or a duplicated child being listed with a grandparent). Some issues appear to have been suppressed in the cognitive data as well. In some GQ cases (jail, shelters) information was not revealed, perhaps because it was sensitive. The contradictory information seen in dependent interviews could also be evidence of a sensitivity to privacy on the part of the respondent.

## **6. Related Census Assessments**

The following assessments, evaluations, and experiments are related to the Effectiveness of Unduplication Evaluation.

- 2010 Census Coverage Followup Assessment
- Evaluation of the Alternative Coverage Followup Questions and Design
- Avoid Followup Evaluation
- Administrative Records Use for Coverage Problems Evaluation
- 2010 Census Field Verification Operational Assessment

## 7. Conclusions and Recommendations

The 2010 Census was the first decennial census where person duplication was identified to be a problem in advance of the census and efforts were made to research and address it during the census. Leading up to the 2010 Census, the Census Bureau tested and implemented the most extensive efforts to date with the goal of resolving duplication. An overview of results from these efforts is presented in this concluding section.

### *Characteristics and Magnitude of Duplication*

From all questionnaires in scope for the Duplicate Person Identification matching process, 7,454,171 person links were identified. HU to HU person links accounted for 88.5 percent of all potential person duplication. HU to GQ links accounted for the remaining 11.5 percent of potential person duplication. Within the HU to GQ links, over half (51.7 percent) were between a HU and student housing, such as a college dormitory.

Of all 7,454,171 person links, 33.5 percent were located within the same block and 10.5 percent were located within surrounding blocks. These duplications are likely caused by address issues on the MAF, form misdeliveries or other housing-level issues. Of the 7,454,171 person links, 28.0 percent were located within the same county, 17.5 percent were located within the same state and 10.5 percent were located across state lines. These duplications are likely caused by complex living situations.

In the majority of HU to HU response links (58.5 percent), only one person was identified as being duplicated between the two HUs. Two people were identified as being duplicated between HUs in 23.6 percent of HU to HU response links.

If two HU responses provided the same phone number on their return, that increased the confidence of persons across each response being a match. Almost one-quarter (24.6 percent) of all HU to HU response links had a matching phone number.

The overcount question on the initial census enumeration was intended to flag possible erroneous enumerations. However, 41.3 percent of suspected HU to GQ duplicates did not mark the overcount question on the HU return and 58.7 percent of suspected HU to HU duplicates did not mark the overcount question on either HU return.

### *CFU Results*

To resolve person-level duplication, more information on living situations is needed than was provided on the initial questionnaire. Mid-decade testing showed that the existing Coverage Followup interview was not successful in resolving duplication, so housing units with persons identified as potential duplicates were not universally eligible for followup as part of the 2010 Census. The success of the Coverage Followup interview was determined by investigating how often respondents mentioned a complex living situation for a potentially duplicated person and how often a potentially duplicated person was removed from a HU roster based on the interview

results, thus resolving the duplication. A sample of cases was selected to be interviewed in 2010 to assess how successful the Coverage Followup interview would be in a true decennial Census environment. There were 469,768 response links sent to Coverage Followup as part of the 2010 Census. A Coverage Followup interview was completed with at least one HU in 74.7 percent of the links. Those 350,757 response links represent 424,806 unweighted person links and 2,209,562 weighted person links.

Only 58.7 percent of person links had at least one side mention a complex living situation in the Coverage Followup interview. Some unduplication cases that were sent to Coverage Followup had marked the overcount question on the initial return, indicating a possible erroneous enumeration, while other unduplication cases sent to Coverage Followup had not marked the overcount question. Person links that had marked the overcount college box were the most likely to mention a complex living situation in the Coverage Followup interview. Person links that had not marked an overcount box on either side of the link were less likely to mention a complex living situation in Coverage Followup than links that had marked the overcount box. For instance, for within county links between two HUs when neither had marked the overcount box, a complex living situation was mentioned in Coverage Followup only 20.2 percent of the time. However, 85.6 percent of links between two HUs where at least one side had marked the college overcount category then mentioned a complex living situation in Coverage Followup.

Of the person links that mentioned a complex living situation in the Coverage Followup interview, 53.2 percent removed someone from the roster. This figure was higher for links between a HU and a GQ; such links deleted someone 73.7 percent of the time when a complex living situation had been mentioned. For all person links (regardless of whether they mentioned a complex living situation or not), the Coverage Followup interview was successful in resolving duplication 31.2 percent of the time. Of the links between two HUs where an indication had been provided on initial census return of a seasonal or second home, 44.0 percent of the links deleted someone as a result of Coverage Followup. That rate increased to 57.8 percent of such links that mentioned a complex living situation in the followup interview. This result is promising for resolving duplication between seasonal residences in a true decennial Census environment.

However, as mentioned earlier, the majority of duplication comes from persons who do not mark an overcount reason on the initial census questionnaire. For person links that were within the same county and neither side flagged the overcount question initially, only 7.0 percent then deleted someone as a result of Coverage Followup. For person links within the same state, 10.1 percent deleted someone as a result of Coverage Followup and 6.9 percent of person links across state lines deleted someone as a result of Coverage Followup. These cases have been and will continue to be the most difficult to resolve.

Overall, the 2010 Coverage Followup operation resolved duplication in 31.2 percent of the person links sent to CFU. The results did show that the CFU interview was able to resolve duplication at a higher rate when at least one of the response links was identified as having a certain overcount reason associated with it, such as college, nursing home, seasonal and military cases. It was also more successful in resolving HU to GQ duplication. However, since the majority of duplication cases do not provide an indication of the living situation, we will likely



need to attempt new followup methods or determine another way to process these duplicate persons without a followup.

### *Experimental results*

From the sample of CFU cases that had duplicated persons on them and completed the experimental series of questions at the end of the CFU interview (Mod Q), most people (57.0 percent) provided an open-ended response as to why they marked the overcount box for that person. The coded results of those responses were not available at the time of this evaluation, but once available could provide additional information about why duplication may occur and not get resolved in CFU.

### *Qualitative and Cognitive Interviews*

Fifty qualitative and 226 cognitive interviews were conducted in 2010 and 2011 to gather more information about the causes of duplication and ways to resolve duplication. Interviews were only conducted with cases that had not marked an overcount category on either initial census questionnaire.

The 2010 experimental TCFU questionnaire tailored followup questions (such as when phone numbers matched, or by the type of GQ in which a duplicate was enumerated). That approach showed promise; most notably, 97.6 percent of phone match cases were verified as duplicates using the tailored TCFU questionnaire. For non-phone match cases, the TCFU interview was particularly unsuccessful with verifying the suspected duplication of an adult (46.2 percent verified) or senior citizen (45.8 percent verified) from a partial household match. In contrast, non-phone matches between whole households of adults were verified 82.4 percent of the time. The cognitive interviews classified 20.0 percent of all non-phone match HU-HU cases (and 13.6 percent of all HU-GQ cases) as being false matches or suspected false matches, indicating that additional improvements could be made to the matching process.

The most common reason for duplication within the cognitive interviews with HU-HU cases was moving, which explained 31.6 percent of all verified duplication cases. Having a second property was the second most common reason for HU-HU cases, followed by staying at a relative's house (outside of custody arrangements).

To confirm the duplication, cognitive interview respondents were asked for the address of other places they stayed. Respondents who had moved or owned a second property were able to provide a complete address of the second place more often than respondents with other reasons for providing an address. For instance, 90.5 percent of the 116 addresses affiliated with a move were considered complete addresses, 72.2 percent of the addresses identified as seasonal homes were considered complete addresses, and all 17 of the addresses otherwise identified as a second property owned by the respondent were considered complete addresses. Complete address information of these other HUs could be useful for automation of the verification process, if captured and fully utilized. In contrast, addresses affiliated with a significant other were considered complete only 21.4 percent of the time, and addresses identified as being a son or daughter's address were complete 42.9 percent of the time. In many cases when a respondent

did not know the complete address of a place, they offered neighborhood information, cross streets, or driving directions as a substitute. Although this information was often adequate to verify duplication in this study, it may not be sufficient for automation of the verification procedures.

Verifying a suspected duplicate (that it is indeed the same person listed on two different census questionnaires) is only the first step; resolving the duplication can be much more difficult. There were seven cases in the cognitive and qualitative interviews where an interviewer spoke with a respondent from each side of the duplicate link. Within these pairs of interviews, the cases are verified in that the address of the other place was revealed in both interviews, and the identity of the duplicated person was agreed upon. Beyond that, however, great differences emerged in the account of the time spent in the two places, despite the apparent openness of the respondents in these situations. Relying on both of these reports jointly, it would not have been possible to determine the proper enumeration for the duplicated person in six of the seven situations.

### *Summary*

As shown in this report, duplication is a problem in the decennial census. Duplication occurred for a variety of reasons in the 2010 Census; complicated living situations that led to a person being enumerated on more than one form, potential duplication of addresses on the Census Bureau's master address file that led to the same household being enumerated twice, respondents and enumerators who did not heed the residence rule instructions or pay attention to April 1 residency, or misdelivery of questionnaires to the intended address. Some of these causes could be addressed internally by the Census Bureau to minimize the occurrence of duplication. Additionally, the computer-matching process that identified duplicates in the 2010 Census is exceptional but can be improved upon; false matches were observed in the cognitive interviewing and the current policy of setting high cutoffs inevitably fails to identify some true matches.

This report was not intended to study address-level duplication with the attention it deserves; that issue merits further investigation. As for person-level duplication, once it was identified in the 2010 Census, there was little chance of resolving the duplication. Only select cases were resolved in the CFU interview, which followed the trend of historical data showing a low success rate in CFU for cases involving duplication. Relying on a followup interview has drawbacks; telephone contact has to first be made, cooperation has to be attained with a respondent, and sufficient information has to then be elicited in a circumspect manner in order to understand the living situation. The data in this report show some bright spots for improving the resolution of duplication in the future however. Moving forward towards 2020, we need to first work on preventing duplication. Given that some duplication is inevitable, we should also consider how to improve the identification of duplicates, as well as what situations can be resolved, and ultimately determine how to resolve more duplication. We provide some recommendations below for how to more effectively unduplicate the census in 2020.

## Recommendations

In order to minimize duplication in future censuses and improve the effectiveness of unduplication efforts that do occur in future censuses, we recommend the following:

- **Gain more knowledge on duplication in the 2010 Census.**
  - **Repeat the analysis presented in Section 5.1 of this report using the complete universe of census returns.** The additional analysis would give us a better picture of the entire universe of duplicates and shed new light on certain situations, (such as GQ-GQ links and duplication from Be Counted forms) that have never been analyzed.
  - **Research large clusters (links involving more than two MAFIDs).** This evaluation only compared links of two addresses, but it is possible for a person to be duplicated between three or more addresses (such as a child of divorced parents who is counted by both parents and also counted in college). Additional research of duplicate links could identify how often large clusters occurred in 2010 and how such situations could be resolved.
  - **Study address-level duplication more, including a further analysis of the Field Verification results from 2010.** A sample of within block and within surrounding block cases were contacted in the Field Verification operation in 2010. The results from that operation do not discuss cases as a linked pair, but only as individual observations. Additional research could also be done to determine if any addresses had been marked for removal from the MAF by a NRFU enumerator but retained on the MAF due to a VDC enumerator or for another reason.
  - **Conduct additional analysis with CFU address data.** In the CFU interview, when a respondent mentioned a complex living situation, they were then asked to provide the address of that place. That address information was collected and sent to GEO to be geocoded. When possible, a MAFID was assigned to the address of that other place. This evaluation did not look at any address information from CFU, but analysis could be conducted with that data, including how much address information respondents were able and willing to provide in CFU and how often the address they gave was the same one where we thought they were duplicated.
  - **Conduct additional research using detailed CFU responses when interviews were completed with both sides of a link.** This would examine how the stories compared, since the decision in 2010 about whether to keep or remove a person from a roster was done while only assessing the CFU results for each side, without comparing responses from both sides of a link (if both sides had been interviewed).
  - **Research the open-ended answers to the first question in Mod Q about why the person was living somewhere else, the usefulness of the address that was collected in Mod Q, and conduct analysis of Mod Q results using linked data.** These results could provide more information on why the CFU interview was not successful with some duplication cases.
- **Attempt to reduce false matches and capture more true matches in future matching.**

- **Explore a way to include relationship status and family composition in the probabilistic weighting of a match, especially when only one person in a household is thought to be duplicated.** We suspect this could reduce the number of false matches made during the matching process, as it was a key component during the clerical review of qualitative interviews. For example, if a child was counted at two housing units with biological fathers at both units, they were less likely to be considered a true match.
- **Consider clarifications to the current categorization of geographic proximity (within block, etc.) or consider using additional variables to describe the actual distance that two addresses are from each other (such as latitude and longitude, or ZIP codes).** When determining the geographic distance of a link (within block, etc.), the current matching process does a hard match comparison of the MAF-provided geocode information for one ID (and an auxiliary file of surrounding blocks) to that same information available for the other ID. This type of comparison does not take into consideration the spatial distance between each ID that may be more useful or relevant. For example, the responses from a link found within the same county could be ten miles apart while the responses from a link found to be within the same state could be ten miles apart (they just happen to be in different counties). Therefore, even though they are spatially the same distance apart, the fact they are currently treated as within county and within state has implications for how small variations or inconsistencies in the comparison data are handled. We recommend researching ways to improve the classification of geographic distance such as through the use of longitude and latitude, ZIP codes, or maybe the addition of surrounding county and state files.
- **Consider the implications and possibilities of automation for identification and resolution of duplicates in the 2020 Census.**
  - **Collect alternate address information when respondents positively indicate they have another place where they live or stay.** Address information would be more useful if it could be collected at the time of the initial census enumeration, instead of weeks or months later in a subsequent contact. Given the results from the TCFU interview where movers were common and apparently cooperative, as well as results from the CFU interview showing seasonal residences to be common and those respondents to be relatively cooperative, we suspect that some living situations would result in respondents who are able and willing to provide additional information on their patterns of residency and locations if we asked for it. Some of this research is also happening now as part of the study known as Cognitive Testing of Roster, Coverage, and Address Questions for the 2020 Census (Childs, 2011). This will also be discussed in the 2010 Census Avoid Followup Evaluation (Jackson and Wechter, 2012).
  - **Once collected, utilize alternate address information to identify duplicates.** One option would be for GEO to geocode all alternative addresses, and then specifically check returns associated with that geocoded MAFID for duplication. If the respondent did not provide enough alternative address information to be

geocoded however, it should be investigated if the partial address information could be incorporated in the matching algorithm to identify duplicates.

- **Seek to improve current practices that influence duplication.**
  - **Continue to research improvements to the overcount question and probes, to capture as many erroneous enumerations as possible.** The overcount question has been modified over the years and can continue to be improved through cognitive testing and followup field work.
  - **Continue to refine the targeted followup approach and utilize existing data about a duplication to make any followup contacts more efficient and successful.** The overcount probes will never be able to capture every erroneous enumeration, so some subsequent followup will be needed. The approach used in the TCFU instrument (where questions were tailored based on age of the duplicate, presence of a phone match, or the type of GQ if applicable) showed promise.
  - **Continue to improve the address list development process to minimize address duplication.** The rate of duplication seen within-block was a significant portion of the total duplication; such links are thought to be the result of address duplication on the MAF. There are various Address List Development steps and decisions that take place during the census operations that may contribute to housing unit duplication on the final census address list (for instance, enumerators are allowed to add addresses in field operations and local governments can submit addresses for review in the LUCA operation). Unduplication of addresses relies on exact matching of addresses collected across operations that occur at different times and with incomplete knowledge of other operational updates. Additionally, the emphasis during enumeration operations is on enumerating people, as compared to resolving address statuses. The interaction of these factors can contribute to duplicated addresses and while improvements have been made to the address list development process since 2000, we recommend continued attention be paid to this process.
  - **Improve the identification of GQs as GQs.** Due to the large amount of within-block duplication seen in HU-GQ links, it is suspected that some physical structures could exist on the MAF twice, as both a HU record and a GQ record (group homes and other non-institutional facilities can be especially difficult to identify). This would result in one enumeration as a HU and one enumeration as a GQ, creating duplication. We recommend exploring ways to minimize this problem from occurring, and also considering ways to identify and resolve it when it does occur.
  - **Expand GEO's matching process to utilize building and GQ names when matching addresses to the MAF.** It was observed during the TCFU interviews that respondents could not provide full city-style addresses for GQs or certain HUs where they stayed, but could provide GQ names or cross streets. Currently, such incomplete address information is not able to be matched by GEO to a known MAFID, but was often adequate for TCFU interviewers to establish that

the incomplete address offered by a respondent was the same as the address at which we thought the person to be duplicated.

- **Consider a clerical operation to resolve lingering duplication cases in the future.** From the TCFU study, almost half of the completed cognitive interviews required clerical Census Bureau data review when the interviewer could not ascertain if the address for the second side had been successfully revealed in the interview. The duplicate address was determined to have been revealed in about one-third of the reviewed cases. Census Bureau staff also reviewed several cases where the interview had not gone as expected, and the clerical review uncovered misclassifications of GQs (which were either classified as HUs or as an incorrect type of GQ). If a large percentage of duplication can be resolved electronically, this workload might be manageable.
  - **Communicate with the public about followup efforts.** The TCFU study mailed advance letters to the targeted respondents to prepare them for a subsequent phone contact, which CFU did not do in 2010. This effort was thought to be successful. If the resolution of duplication requires cooperation by respondents in a followup interview, we should improve our efforts to secure their cooperation.
- **Review current Census Bureau policies related to duplication.**
    - **Review the policies for contacting suspected duplicates to ensure that no violations of privacy or confidentiality occur in any new methods of resolving duplication.** Efforts to unduplicate the census are constrained by important Census Bureau policies that protect respondent privacy and the confidentiality of data. Provisional allowances were made for the new procedures and methods used in the TCFU study, including revealing both addresses to Type 1 cases where the phone number matched between sides. As these methods develop, continued attention should be paid to the policies surrounding these issues.
    - **For HU-GQ links, if alternate address information can be collected from the initial returns, processed automatically, and utilized to confirm duplication, then consider resolving the duplication according to the residence rule without any additional contacts to the HU.** The residence rule states that people in certain types of GQs on Census Day should be counted at the GQ.
    - **If alternate address information and information on living patterns can be collected initially for certain HU-HU duplicate situations (such as movers or seasonal residences), then consider resolving the duplication according to the residence rule without any additional contacts to the HU.** The residence rule states that people should be counted at their usual residence, which is the place where they live and sleep most of the time.
    - **Review the implications of counting persons at multiple residences.** The Census residence rule states that we need to count people at their usual residence, which is where they live and sleep most of the time; thus, we are required to attempt to resolve duplication. Some living situations that lead to duplication are difficult to resolve, such as child custody cases. If we allow persons to be partially counted at multiple places, then this may be more reflective of how persons live and a followup interview might not be necessary, which would result

in a cost savings. We recommend researching what would have happened during apportionment and redistricting if we had not attempted to resolve duplication and allowed persons identified as duplicates to be counted partially at each living quarter they were enumerated at. As part of this research, the legal and political implication of such a decision should be considered.

- **Consider how Administrative Records could be utilized either to confirm a suspected duplication or aid in the resolution of where a duplicated person should be counted.** The matching process takes a conservative approach to the identification of duplicate persons in order to minimize the number of false matches, which also inherently then fails to identify a number of true matches. We recommend researching the use of Administrative Records as a supplemental data source in the identification of duplicate persons. We also recommend researching the use of Administrative Records to learn if there is information that can be used to determine which one place to count a person. If it can be shown that Administrative Records are useful in the reconciliation of duplication, then there will be a more accurate census and a savings in costs.

We were unable to resolve much duplication in the 2010 Census but significant advancements can be made to greatly reduce duplication in the 2020 Census.

## 8. Acknowledgements

We would like to recognize and thank Maureen Lynch, Susan Frank, Courtney Ford, Leah Marshall, Robin Pennington, Mike Ikeda and Ned Porter. Without their expertise on matching and identifying person links, this report would not be possible.

We would like to recognize and thank Elizabeth Poehler for her continued guidance and support on all unduplication related topics.

We would like to recognize and thank Larry Cahoon who provided valuable guidance on the methods used to produce the weights used in this report.

We would like to thank the staff at RTI International and RSS for the successful and insightful qualitative portion of this evaluation. We would also like to thank Jennifer Childs for coordinating this effort.

We would like to thank Kelly Govern, Julia Coombs, Timothy Stewart and Danquan Prunty for their help with CFU and Mod Q portions of this report.

We would like to thank Elizabeth Poehler, Geoff Jackson, and Kelly Govern for providing comments on an early version of this evaluation.

## 9. References

- Barrett, Diane (2010). "Specifications for the 2010 Census Data Capture Audit and Resolution-Version 3," DSSD 2010 Decennial Census Memorandum Series #C-01R4, U.S. Census Bureau, January 27, 2010.
- Beaghen, Michael and Byrne, Rose (2002). "Clerical Review of Census Duplicates," Decennial Statistical Studies Division A.C.E. Revision II Memorandum Series #PP-43, December 31, 2002.
- Bean, Susanne L. and Bauder, D. Mark (2002). "Census and Administrative Records Duplication Study," Decennial Statistical Studies Division A.C.E. Revision II Memorandum Series #PP-44, December 31, 2002.
- Blough, Laura (2010). "2010 Census Coverage Followup Assessment Study Plan." 2010 Census Planning Memoranda Series No. 80, U.S. Census Bureau, May 17, 2010.
- Childs, Jennifer Hunter, Fernandez, Leticia, Clifton, Matt, and Myers, Mikelyn (2009). "CFU CPEX Experimental Question Cognitive Testing: Undercount, Overcount, and Duplicate Experimental Question Sequences." SRD Statistics Research Report Series #2009-17, U.S. Census Bureau, November 16, 2009.
- Childs, Jennifer Hunter, Sorokin, Anissa, and Jurgenson, Nathan (2011). "2009 Targeted Coverage Follow-up (TCFU) Cognitive Testing." SRD Statistics Research Report Series #2011-06, U.S. Census Bureau, May 31, 2011.
- Childs, Jennifer Hunter (2011). "Study Plan for Cognitive Testing of Roster, Coverage, and Address Questions for the 2020 Census." September 2011. DRAFT.
- Fay, R.E. (2001). "The 2000 Housing Unit Duplication Operations and Their Effect on the Accuracy of the Population Count," *2001 Proceedings of the Joint Statistical Meetings on CD-ROM*, American Statistical Association, Alexandria, VA.
- Fay, R.E. (2002). "Probabilistic Models for Detecting Census Person Duplication," *2002 Proceedings of the Joint Statistical Meetings on CD-ROM*, American Statistical Association, Alexandria, VA, pp. 969-974.
- Fay, R.E. (2003). "Probabilistic Models for Detecting Census Duplication at the Person and Household Levels," *2003 Proceedings of the Joint Statistical Meetings on CD-ROM*, American Statistical Association, Alexandria, VA, pp. 1391-1398.
- Fay, R.E. (2004). "An Analysis of Person Duplication in Census 2000." *2004 Proceedings of the Joint Statistical Meetings on CD-ROM*, American Statistical Association, Alexandria, VA, pp. 3478-3485.



- Feldpausch, R (2001). "Census Person Duplication and Corresponding A.C.E. Enumeration Status," Executive Steering Committee on Accuracy and Coverage Evaluation Policy II Report 6, U.S. Census Bureau. October 13, 2001.
- Frank, Sue, Ikeda, Michael, and Porter, Edward (2011a). "2010 Decennial Census Coverage Followup and Census Coverage Measurement Person Matching Parameter Software Requirements Specification," DSSD Decennial Census 2010 Memorandum Series #I-02R1, U.S. Census Bureau, January 12, 2011.
- Frank, Sue, Ikeda, Michael, and Porter, Edward (2011b). "2010 Decennial Census Coverage Followup and Census Coverage Measurement Match Modeling Software Requirements Specification," DSSD Decennial Census 2010 Memorandum Series #I-03R1, U.S. Census Bureau, January 12, 2011.
- Govern, Kelly, Kostanich, Martine, and Heimel, Sarah (2009). "2008 Coverage Followup Assessment." DSSD 2008 Census Dress Rehearsal Memorandum Series #O-18, U.S. Census Bureau, June 3, 2009.
- Govern, Kelly, Coombs, Julia, and Glorioso, Robert (2012). "2010 Census Coverage Followup Assessment." 2010 Census Planning Memoranda Series, No. 197, U.S. Census Bureau, May 29, 2012.
- IBM, "DRIS CFU Application Design Document", Version 38, November 2007
- Heimel, Sarah (2010). "2010 Census Study Plan for the Effectiveness of Unduplication Evaluation." 2010 Census Planning Memorandum Series #110, U.S. Census Bureau, December 8, 2010.
- Ikeda, Michael and Porter, Edward (2007). "Initial Results from a Nationwide BigMatch Matching of 2000 Census Data." SRD Statistics Research Report Series #2007-22, U.S. Census Bureau, December 29, 2007.
- Ikeda, Michael and Porter, Edward (2008). "Additional Results from a Nationwide Matching of 2000 Census Data." SRD Statistics Research Report Series #2008-2, U.S. Census Bureau, March 5, 2008.
- Jackson, Geoffrey, and Heimel, Sarah (2010). "2010 Census Avoid Followup Evaluation Study Plan." 2010 Census Planning Memorandum Series #108, U.S. Census Bureau, December 3, 2010.
- Jackson, Geoffrey and Wechter, Keith (2012). "2010 Census Avoid Followup Evaluation." 2010 Census Planning Memorandum Series #236, U.S. Census Bureau, September 7, 2012.
- Kostanich, Martine (2009). "Identification of the 2010 Coverage Followup Eligible Universe and Selection Requirements." DSSD 2010 Decennial Census Memorandum Series #I-06, U.S. Census Bureau, April 30, 2009.

- Krejsa, Elizabeth, Linse, Kyra, Kostanich, Martine, Heimel, Sarah, Marshall, Leah, Banz, Edward, and King, Ryan (2007), "2006 Census Test Evaluation: Evaluation 2: Coverage Improvement," DSSD 2006 Census Test Memorandum Series #F-05, U.S. Census Bureau, September 24, 2007.
- Lamas, Enrique (2009), "Residence Rule and Residence Situations for the 2010 Census." 2010 Census Information Memoranda Series No. 35, U.S. Census Bureau, April 24, 2009.
- Lynch, Maureen (2009a). "2010 Decennial Census Coverage Followup Duplicate Person Identification Software Requirements Specification," DSSD Decennial Census 2010 Memorandum Series #I-01, U.S. Census Bureau, May 29, 2009.
- Marshall, Leah (2007). "Analysis Plan for Unduplication Cases in the 2006 Census Test Evaluation of Coverage Improvement." DSSD 2006 Census Test Memorandum Series #I-12, U.S. Census Bureau, June 18, 2007.
- Marshall, Leah (2008a). DSEP Issue Paper. September 4, 2008.
- Marshall, Leah (2008b). "Potential Duplicates in the Census – Methodology and Selection of Cases for Followup." Paper presented at the American Association for Public Opinion Research Annual Conference, New Orleans, LA, May 15, 2008.
- Marshall, Leah (2009a). "Customer Requirements for Surrounding Block Files for 2010 Decennial Census Unduplication Case Selection." DSSD Decennial Census 2010 Memorandum Series #I-04, U.S. Census Bureau, January 6, 2009.
- Marshall, Leah (2009b). "2010 Decennial Census Unduplication Universe Selection Software Requirements Specification for Coverage Followup, Field Verification, and Evaluation Processing." DSSD 2010 Decennial Census Memorandum Series #I-05, U.S. Census Bureau, July 7, 2009.
- Marshall, Leah (2009c). DSEP Issue Paper. September 18, 2009.
- McPhillips, Frank (2012). "2010 Census Field Verification Assessment Report," 2010 Census Planning Memoranda Series No. 162 (Reissue), February 23, 2012.
- Milligan, Mitch (2010). "2010 Field Verification Processing Assessment Study Plan." DRAFT.
- Mule, Thomas (2001). "Executive Steering Committee on Accuracy and Coverage Evaluation Policy II Report Number 20: Person Duplication in Census 2000." DSSD Census 2000 Procedures and Operations Memorandum Series Q. U.S. Census Bureau, October 11, 2001.
- Mule, Thomas (2002). "A.C.E. Revision II Results: Further Study of Person Duplication." DSSD A.C.E Revision II Memorandum Series #PP-51. U.S. Census Bureau, December 31, 2002.

- Mule, Thomas (2003). "Estimate of Census 2000 Person Duplication by Type of Returns." 2004 Census Test Memorandum Series # G-5, U.S. Census Bureau, June 27, 2003.
- Nash, Fay (2000a). "Overview of the Duplicate Housing Unit Operations." Census 2000 Informational Memorandum No. 78, U.S. Census Bureau, November 7, 2000.
- Nash, Fay (2000b). "Results of Reinstatement Rules for the Housing Unit Duplication Operations." Census 2000 Informational Memorandum Mo. 82, U.S. Census Bureau, November 21, 2000.
- Pennington, Robin (2005a). "Evaluation of Person Duplication." 2004 Census Test Evaluation 8. U.S. Census Bureau, September 28, 2005.
- Pennington, Robin A. (2005b). "Unduplication of Persons and Housing Units in the 2004 Census Test." In *2005 JSM Proceedings*, Social Statistics Section. Alexandria, VA: American Statistical Association.
- Peytcheva, Emilia, Sha, Mandy, Gerber, Eleanor, Cook, Sarah, Schoua-Glusberg, Alisú, King, Tiffany, Kenward, Katherine (2011). "Qualitative Interviewing with Suspected Duplicates and Cognitive Testing of the Targeted Coverage Follow-up (TCFU) Interview." Prepared for the U.S. Census Bureau by RTI International and Research Support Services. September, 2011.
- Poehler, Elizabeth Krejsa (2010a). "Decision on Contingency Workloads for the 2010 Coverage Followup Operation." 2010 Decennial Census Program Decision Memorandum Series Number 31, U.S. Census Bureau, April 1, 2010.
- Poehler, Elizabeth Krejsa (2010b). "Decision on the Workloads for the 2010 Coverage Followup and Field Verification Operations." 2010 Decennial Census Program Decision Memorandum Series Number 28 (reissue), U.S. Census Bureau, April 1, 2010.
- Sheppard, Dave (2003). "Coverage Edit Followup," Census Evaluation I.1, U.S. Census Bureau, July 29, 2003.
- Smith, Damon (2004). "Long Distance Duplicate Telephone Followup Study." DSSD 2003 Memorandum Series Chapter #A-03. July 8, 2004.
- Stewart, Timothy D. (2010). "2010 Census Evaluations, Experiments, and Assessments Study Plan – Alternative Coverage Followup Questions and Design." 2010 Census Planning Memorandum Series #94, U.S. Census Bureau, August 18, 2010.
- Yancey, William E. and Winkler, William E. (2002). "Record Linkage Software: User Documentation." U.S. Census Bureau, October 10, 2002.
- Yancey, William (2007), "BigMatch: A Program for Extracting Probable Matches from a Large File," SRD Research Report Series #2007-01, U.S. Census Bureau, June 15, 2007.

## APPENDIX A – 2010 Census Residence Rule and Residence Situations

### Residence Rule

The residence rule is used for determining where people should be counted (which means tabulated) in the 2010 Census.

**Residence Rule:** Count people at their usual residence, which is the place where they live and sleep most of the time. People in certain types of group quarters (GQ) on Census Day should be counted at the GQ. These GQ types are listed in the box below. People who do not have a usual residence or cannot determine a usual residence, and who are not in one of the GQs types listed below, should be counted where they are on Census Day.

### Group Quarters in Which All People Are Counted at the Group Quarters

[For more details on these GQs and for the other GQ types in which the residents may be counted at their usual residence other than the GQ, see Situation 13 on pages 5-7]

<u>Situation</u>	<u>GQ type</u>
13.1	Correctional residential facilities
13.2	Federal detention centers
13.3	Federal and state prisons
13.4	Local jails and other municipal confinement facilities
13.5	Group homes intended for adults (non-correctional)
13.7	Hospitals - count only patients who have no usual home elsewhere
13.9	Mental (psychiatric) hospitals and psychiatric units for long-term non-acute care in other hospitals
13.10	Nursing facilities/skilled-nursing facilities
13.11	Correctional facilities intended for juveniles
13.12	Group homes for juveniles (non-correctional)
13.13	Residential treatment centers for juveniles (non-correctional)
13.14	Military disciplinary barracks and jails
13.15	Military quarters (excluding military vessels)
13.17	Military treatment facilities - count only assigned active duty patients
13.18	College/university student housing
13.19	Residential schools for people with disabilities
13.20	Domestic violence shelters
13.21	Emergency and transitional shelters (with sleeping facilities) for people experiencing homelessness
13.22 (part)	Targeted non-sheltered outdoor locations

**The following section shows how the Census Bureau applies the  
2010 residence rule to various residence situations.**

<b>Application of the 2010 Residence Rule to Residence Situations</b>		
	<b>Residence situation</b>	<b>Counted at -</b>
<b>1. PEOPLE AWAY FROM THEIR USUAL RESIDENCE ON CENSUS DAY</b>		
	1.1 People away from their usual residence on Census Day, such as on a vacation or business trip, visiting, traveling outside the U.S., or working elsewhere without a usual residence there (for example, as a truck driver or traveling salesperson)	Residence where they live and sleep most of the time
	1.2 Patients in general or Veterans Affairs hospitals (except psychiatric units) on Census Day, including newborn babies still in the hospital on Census Day	Residence where they live and sleep most of the time. For newborn babies, residence where they will live and sleep most of the time
<b>2. VISITORS ON CENSUS DAY</b>		
	2.1 Visitors on Census Day who will return to their usual residence (For foreign visitors, see 10.3.)	Residence where they live and sleep most of the time
<b>3. PEOPLE WHO LIVE IN MORE THAN ONE PLACE</b>		
	3.1 People living away most of the time while working, such as people who live at a residence close to where they work and return regularly to another residence	Residence where they live and sleep most of the time. If time is equally divided, residence where they are staying on Census Day. If there is no residence where they live most of the time, residence is where they live and sleep more than anywhere else
	3.2 People who live at two or more residences (during the week, month, or year), such as people who travel seasonally between residences (for example, snowbirds)	Residence where they live and sleep most of the time. If time is equally divided, residence is where they are staying on Census Day. If there is no residence where they live most of the time, residence is where they live and sleep more than anywhere else
	3.3 Children in shared custody or other arrangements who live at more than one residence	Residence where they live and sleep most of the time. If time is equally divided, residence where they are staying on Census Day
<b>4. STUDENTS</b>		
	4.1 Boarding school students living away from their parental home while attending boarding school below the college level, including Bureau of Indian Affairs boarding schools	Their parental home
	4.2 College students living at their parental home while attending college	Their parental home

	4.3 College students living away from their parental home while attending college in the U.S. (on-campus or off-campus)	On-campus or off-campus residence where they live and sleep most of the time
	4.4 College students living away from their parental home while attending college in the U.S. (on-campus or off-campus) but staying at their parental home while on break or vacation	On-campus or off-campus residence where they live and sleep most of the time
	4.5 U.S. college students living outside the U.S. while attending college outside the U.S.	Not counted in the census
	4.6 Foreign students living in the U.S. while attending college in the U.S. (on-campus or off-campus)	On-campus or off-campus residence where they live and sleep most of the time
<b>5. MOVERS ON CENSUS DAY</b>		
	5.1 People who move into a residence on Census Day who have not been listed on a questionnaire for any residence	Residence they move into on Census Day
	5.2 People who move out of a residence on Census Day and have not moved into a new residence on Census Day and who have not been listed on a questionnaire for any residence	Residence they move out of on Census Day
	5.3 People who move out of a residence or move into a residence on Census Day who have been listed on a questionnaire for any residence	Not counted again
<b>6. PEOPLE WHO ARE BORN OR DIE ON CENSUS DAY</b>		
	6.1 Babies born on or before 11:59:59 p.m. on Census Day	Residence where they will live and sleep most of the time
	6.2 Babies born after 11:59:59 p.m. on Census Day	Not counted in the census
	6.3 People who die before 12:00:00 a.m. on Census Day (that is, people who die at 11:59:59 p.m. on March 31 or earlier)	Not counted in the census
<b>7. NONRELATIVES OF THE HOUSEHOLDER</b>		
	7.1 Roomers or boarders	Residence where they live and sleep most of the time
	7.2 Housemates or roommates	Residence where they live and sleep most of the time
	7.3 Unmarried partners	Residence where they live and sleep most of the time
	7.4 Foster children or foster adults	Residence where they live and sleep most of the time
	7.5 Live-in employees, such as caregivers or domestic workers	Residence where they live and sleep most of the time

<b>8. U.S. MILITARY PERSONNEL</b>		
	8.1 U.S. military personnel living in barracks in the U.S.	Barracks
	8.2 U.S. military personnel living on base or off base not in barracks in the U.S.	Residence where they live and sleep most of the time
	8.3 U.S. military personnel on U.S. military vessels with a U.S. homeport	Onshore U.S. residence where they live and sleep most of the time; otherwise at their vessel's homeport
	8.4 U.S. military personnel living on or off a military installation outside the U.S., including dependents living with them	Not included by the stateside enumeration <sup>52</sup>
	8.5 U.S. military personnel on U.S. military vessels with a homeport outside the U.S.	Not included by the stateside enumeration <sup>53</sup>
<b>9. MERCHANT MARINE PERSONNEL ON U.S. FLAG MARITIME/MERCHANT VESSELS</b>		
	9.1 Crews of U.S. flag maritime/merchant vessels on Census Day docked in a U.S. port or sailing from one U.S. port to another U.S. port	Onshore U.S. residence where they live and sleep most of the time; otherwise at their vessel. If the vessel is docked in a U.S. port, it is counted at the port. If the vessel is sailing from one U.S. port to another U.S. port, it is counted at the port of departure.
	9.2 Crews of U.S. flag maritime/merchant vessels engaged in U.S. inland waterway transportation on Census Day	Onshore residence where they live and sleep most of the time
	9.3 Crews of U.S. flag maritime/merchant vessels on Census Day docked in a foreign port, sailing from one foreign port to another foreign port, sailing from a U.S. port to a foreign port, or sailing from a foreign port to a U.S. port	Not enumerated or counted in the census
<b>10. FOREIGN CITIZENS IN THE U.S.</b>		
	10.1 Citizens of foreign countries living in the U.S.	U.S. residence where they live and sleep most of the time
	10.2 Citizens of foreign countries living in the U.S. who are members of the diplomatic community	Embassy, consulate, United Nations' facility, or other residences where diplomats live. They have the right to refuse to provide any or all information
	10.3 Citizens of foreign countries visiting the U.S., such as on a vacation or business trip	Not enumerated or counted in the census

<sup>52</sup>Included in state counts for apportionment purposes based on *Franklin v. Massachusetts*, 112 S. Ct. 2767 (1992).

<sup>53</sup>Included in state counts for apportionment purposes based on *Franklin v. Massachusetts*, 112 S. Ct. 2767 (1992).

<b>11. U.S. CITIZENS AND THEIR DEPENDENTS LIVING OUTSIDE THE U.S.</b>		
	11.1 U.S. citizens living outside the U.S. and employed as civilians by the U.S. Government, including dependents living with them	Not included by the stateside enumeration <sup>54</sup>
	11.2 U.S. citizens living outside the U.S. and not employed by the U.S. Government and not dependents as described in 8.4 and 11.1	Not enumerated or counted in the census
<b>12. PEOPLE AT TRANSITORY LOCATIONS</b>		
	12.1 People at transitory locations such as recreational vehicle (RV) parks, campgrounds, hotels and motels (including those on military sites), hostels, marinas, racetracks, circuses, or carnivals	Residence where they live and sleep most of the time. If time is equally divided, residence where they are staying on Census Day. If there is no residence where they live most of the time, residence is where they live and sleep more than anywhere else
<b>13. PEOPLE IN GROUP QUARTERS</b> [2010 GQ code in parentheses. For detailed definitions of GQ types, see the 2010 Census Group Quarters Definitions and Code List, January 30, 2009]		
	<b>People in Correctional Facilities for Adults on Census Day</b>	
	13.1 People in correctional residential facilities on Census Day (105)	GQ [UHE (usual home elsewhere) not allowed]
	13.2. People in federal detention centers on Census Day (101)	GQ (UHE not allowed)
	13.3 People in federal and state prisons on Census Day (102 federal/103 state)	GQ (UHE not allowed)
	13.4 People in local jails and other municipal confinement facilities on Census Day (104)	GQ (UHE not allowed)
	<b>People in Group Homes and Residential Treatment Centers for Adults</b>	
	13.5 People in group homes intended for adults (non-correctional) (801)	GQ (UHE not allowed)
	13.6 People in residential treatment centers for adults (non-correctional) (802)	Residence where they live and sleep most of the time (UHE allowed); otherwise at the GQ
	<b>People in Health Care Facilities</b>	
	13.7 People in hospitals on Census Day who have no usual home elsewhere (402)	GQ (UHE not allowed)
	13.8 People staying in in-patient hospice facilities on Census Day (403)	Residence where they live and sleep most of the time (UHE allowed); otherwise at the GQ

<sup>54</sup>Included in state counts for apportionment purposes based on *Franklin v. Massachusetts*, 112 S. Ct. 2767 (1992).



	13.9 People in mental (psychiatric) hospitals and psychiatric units for long-term non-acute care in other hospitals on Census Day (401)	GQ (UHE not allowed)
	13.10 People in nursing facilities/skilled nursing facilities on Census Day (301)	GQ (UHE not allowed)
	<b>People in Juvenile Facilities</b>	
	13.11 People in correctional facilities intended for juveniles on Census Day (203)	GQ (UHE not allowed)
	13.12 People in group homes for juveniles (non-correctional) on Census Day (201)	GQ (UHE not allowed)
	13.13 People in residential treatment centers for juveniles (non-correctional) on Census Day (202)	GQ (UHE not allowed)
	<b>People in Military Group Quarters</b>	
	13.14 People in military disciplinary barracks and jails (106)	GQ (UHE not allowed)
	13.15 People in military quarters, including barracks (601)	GQ (UHE not allowed)
	13.16 People in military ships (602)	Onshore residence where they live and sleep most of the time (UHE allowed); otherwise at their vessel's homeport
	13.17 People in military treatment facilities with assigned active duty patients (404)	GQ (UHE not allowed)
	<b>People in Residential School-Related Facilities</b>	
	13.18 People in college/university student housing (501)	GQ (UHE not allowed)
	13.19 People in residential schools for people with disabilities on Census Day (405)	GQ (UHE not allowed)
	<b>People in Service-Based Enumeration Facilities</b>	
	13.20 People in domestic violence shelters on Census Day (703)	GQ (UHE not allowed)
	13.21 People in emergency and transitional shelters (with sleeping facilities) on Census Day for people experiencing homelessness (701)	GQ (UHE not allowed)
	13.22 People at soup kitchens (702), regularly scheduled mobile food vans (704), and targeted non-sheltered outdoor locations (706)	<i><b>Soup kitchens and food vans:</b></i> Residence where they live and sleep most of the time (UHE allowed); otherwise at the GQ <i><b>Targeted non-sheltered outdoor locations:</b></i> GQ (UHE not allowed)

	<b>People in Other Facilities</b>	
	13.23 People in living quarters for victims of natural disasters (903)	Residence where they live and sleep most of the time (UHE allowed); otherwise at the GQ
	13.24 Crews on U.S. flag maritime/merchant vessels (900)	Onshore residence where they live and sleep most of the time (UHE allowed); otherwise at their vessel. If the vessel is docked in a U.S. port, it is counted at the port. If the vessel is sailing from one U.S. port to another U.S. port, it is counted at the port of departure.
	13.25 People in religious group quarters (902)	Residence where they live and sleep most of the time (UHE allowed); otherwise at the GQ
	13.26 People in workers' group living quarters and Job Corps Centers (901)	Residence where they live and sleep most of the time (UHE allowed); otherwise at the GQ

## APPENDIX B – Person Counts by Form Type by Wave

The following table shows the availability of person records by form type as they became available on the DRF in 2010, up until July 27<sup>th</sup>. The column labeled ‘Other’ contains all CPEX, TQA, and Fulfillment questionnaires. The column labeled ‘NRFU’ contains questionnaires enumerated during NRFU, VDC, Residual, UE, RUE, RA.

W01 stands for the first wave of data. Waves 8 and 9 had two components each, identified as A and B. Wave 9A was the last wave of person records that were available in time to be matched and possibly sent to CFU. The person records in waves 9B, 10, and 11 were thus not included in the CFU analysis in this report.

July 27, 2010

Person Counts - Form Types							
Wave	UL	Mailback	Replacement	NRFU	Other	GQs	Total Persons
W01	1,372,429	39,945,264	-	4,724	40,254		41,362,671
W02	411,095	35,910,539	207,743	78,976	32,161		36,640,514
W03	194,114	27,969,204	1,543,638	93,239	966,477		30,766,672
W04	124,468	32,606,031	1,585,734	190,661	299,327	130,306	34,936,527
W05	108,983	36,177,763	2,355,185	59,127	247,626	209,730	39,158,394
W06	49,483	24,778,315	4,549,965	94,245	73,917	656	29,546,581
W07	16,221	11,182,227	1,726,262	1,190,056	29,022	431,928	14,575,716
W8A	5,354	1,196,136	267,472	3,838,026	15,576	2,405,902	7,728,466
W8B	4,144	495,362	154,074	10,092,387	7,815	3,570,234	14,324,016
W9A	423	58,322	23,232	3,537,586	852	4,196	3,624,611
W9B	102,828	230,337	74,557	21,967,771	3,308	124,971	22,503,772
W10	308	46,960	14,142	14,565,150	945	97,087	14,724,592
W11	1,034,837	402,835	62,577	4,130,379	413,574	178,124	6,222,326
Totals	3,424,687	210,999,295	12,564,561	59,842,327	2,130,854	7,153,134	296,114,858

Source: DSSD Processing Systems and Development Branch Duplicate Person Identifier Logs